Development and Evaluation of a Method for Measuring Breast Density

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy (PhD) in the Faculty of Medical and Human Sciences

2012

Jennifer Louise Diffey

School of Medicine
Contents

Contents ............................................................................................................................................... 2
Figures and Tables .............................................................................................................................. 7
Abbreviations ..................................................................................................................................... 15
Abstract .............................................................................................................................................. 16
Declaration ......................................................................................................................................... 17
Copyright Statement .......................................................................................................................... 18
Acknowledgements ........................................................................................................................... 19
The Author ......................................................................................................................................... 20
1. Introduction .................................................................................................................................... 22
  1.1 Breast density .............................................................................................................................. 22
  1.2 Aims and objectives .................................................................................................................... 26
    1.2.1 Extend the Manchester Method for volumetric breast density measurement .......... 26
    1.2.2 Validate the new method ......................................................................................................... 26
    1.2.3 Evaluate the feasibility of using the improved method in the breast screening programme . 27
    1.2.4 Obtain information on the distribution of volumetric breast density in the screening population . ............................................................................................................................................................ 27
    1.2.5 Correlate absolute glandular volume and percentage breast density with established breast cancer risk factors and risk ................................................................................................................ 27
    1.2.6 Adapt the method to make it suitable for use in digital mammography ......................... 27
    1.2.7 Adapt the method to measure density by area and examine the association with visually assessed density ............................................................................................................................... 28
  1.3 Thesis outline ............................................................................................................................... 28
2. Background .................................................................................................................................... 30
  2.1 National Health Service Breast Screening Programme (NHSBSP) .......................................... 30
  2.2 Physical principles of mammography ......................................................................................... 33
  2.3 Anatomy of the breast ................................................................................................................. 35
  2.4 Breast density classification schemes ........................................................................................ 38
    2.4.1 Wolfe Classification Scheme ................................................................................................... 38
    2.4.2 Boyd Classification Scheme ..................................................................................................... 40
    2.4.3 BI-RADS Lexicon ..................................................................................................................... 41
    2.4.4 Tabár Patterns .......................................................................................................................... 41
  2.5 Clinical studies of breast density ................................................................................................. 42
  2.6 Possible mechanisms of breast cancer risk associated with breast density ............................. 43
  2.7 Breast cancer risk factors ............................................................................................................ 44
    2.7.1 Age ............................................................................................................................................ 45
    2.7.2 Menarche and menopause ....................................................................................................... 46
    2.7.3 Parity and age at first birth ....................................................................................................... 47
    2.7.4 Weight / Body Mass Index (BMI) ............................................................................................. 48
4.7 Discussion ................................................................................................................................. 114
5. Uncertainties in the Manchester Method ......................................................................................... 117
5.1 Sources of uncertainty .................................................................................................................. 117
5.1.1 Scattered radiation .................................................................................................................. 117
5.1.2 Anode-heel effect ................................................................................................................... 117
5.1.3 Calibration data ..................................................................................................................... 118
5.2 Effect of inaccurate breast thickness on percentage breast density ................................................. 119
5.3 Intra-observer analysis of the Manchester Method ........................................................................ 122
5.4 Inter-observer analysis of the Manchester Method ...................................................................... 125
5.4.1 CC view .................................................................................................................................. 126
5.4.2 MLO view ................................................................................................................................ 130
5.5 Discussion .................................................................................................................................. 131
6. Validation and Comparison ........................................................................................................... 132
6.1 Inter-observer analysis of visually-assessed mammographic density ............................................ 133
6.1.1 Method .................................................................................................................................... 133
6.1.2 Results .................................................................................................................................... 133
6.1.3 Discussion of findings ............................................................................................................. 135
6.2 Correlation of volumetric breast density using the Manchester Method with visually assessed mammographic density ................................................................. 137
6.3 Correlation of area-based breast density using the Manchester Method with visually assessed mammographic density ................................................................................................................... 142
6.3.1 Method .................................................................................................................................... 142
6.3.2 Results .................................................................................................................................... 142
6.3.3 Discussion of findings ............................................................................................................. 146
6.4 Discussion .................................................................................................................................. 147
7. Application of the Manchester Method to the Screening Population ............................................. 148
7.1 Aims of the study ........................................................................................................................ 148
7.2 Method ....................................................................................................................................... 149
7.3 Results ....................................................................................................................................... 151
7.3.1 Consent rate ........................................................................................................................... 151
7.3.2 Descriptive Statistics .............................................................................................................. 151
7.3.2.1 Compressed breast thickness ............................................................................................ 151
7.3.2.2 Breast area .......................................................................................................................... 153
7.3.2.3 Breast volume ...................................................................................................................... 154
7.3.2.4 Breast density ...................................................................................................................... 155
7.3.3 Variation with view .................................................................................................................. 157
7.3.4 Variation of breast density with compressed breast thickness and breast volume ............ 167
7.3.5 Method failures ....................................................................................................................... 168
7.3.6 Comments from invited women ............................................................................................. 173
7.4 Discussion .................................................................................................................................. 175
8. Relationship between Volumetric Breast Density and Breast Cancer Risk Factors

8.1 Study Questionnaires

8.2 Relationship between breast density and breast cancer risk factors

8.2.1 Age at Examination

8.2.2 Weight and Body Mass Index (BMI)

8.2.3 Family history

8.2.4 Ages at menarche and menopause

8.2.5 Parity versus nulli-parity

8.2.6 Ethnicity

8.2.7 Hormone Replacement Therapy (HRT)

8.2.8 Breast cancer

8.2.9 Smoking

8.2.10 Deodorant use

8.2.11 Socioeconomic status

8.3 Correlation of risk factors with area-based density

8.4 Validity of self-reported data

8.4.1 Accuracy of data

8.4.2 Reliability of data

8.4.3 Confidence in responses

8.5 Discussion

9. Adaptation of the Manchester Method to Full Field Digital Mammography

9.1 Digital detector technology

9.1.1 Computed Radiography (CR)

9.1.2 Integrated Digital Detectors (DR)

9.2 The Manchester Method for Full Field Digital Mammography

9.3 Detector stability

9.4 Discussion

10. Conclusions and Future Work

10.1 Meeting the aims and objectives

10.1.1 Extend the Manchester Method for volumetric breast density measurement

10.1.2 Validate the new method

10.1.3 Evaluate the feasibility of using the improved method in the breast screening programme

10.1.4 Obtain information on the distribution of volumetric breast density in the screening population

10.1.5 Correlate absolute glandular volume and percentage breast density with established breast cancer risk factors and risk

10.1.6 Adapt the method to make it suitable for use in digital mammography

10.1.7 Adapt the method to measure density by area and examine the association with visually assessed density
Figures and Tables

Figure 1.1: Boyd categories for mammographic density......................................................... 23
Figure 2.1: Breast Cancer, European Age-Standardised Incidence and Mortality Rates,
UK Females, 1975-2008 ......................................................................................................... 30
Table 2.1: Programme statistics (England), Compiled from the NHSBSP Statistical Bulletin, 2010-
11 ................................................................................................................................................. 31
Figure 2.2: Left: Schematic representation of mammography examination;
Right: typical mammography x-ray unit .................................................................................. 33
Figure 2.3: Subject contrast of malignant tissue and calcification, relative to fibrous tissue. .... 34
Figure 2.4: Unfiltered molybdenum x-ray spectrum (dotted blue line) and effect of adding 0.03mm
molybdenum filter (solid pink line). Generated using IPEM Report 78 Spectrum Processor .... 35
Figure 2.5: Normal anatomy of the breast, lateral view .......................................................... 36
Figure 2.6: Orientation and schematic representation of the cranio-caudal (CC) and medio-lateral
oblique (MLO) projections and images ...................................................................................... 36
Figure 2.7: Digital mammogram images with labelled components of the parenchyma .......... 37
Table 2.2: Established risk factors for breast cancer ............................................................... 44
Table 2.3: Age-specific risk of breast cancer ........................................................................... 45
Figure 2.8: Pike model of breast tissue ageing (A) and log-log plot of age specific incidence of
breast cancer in the USA, 1969 – 71, with fitted curve from Pike model (B) ......................... 46
Figure 2.9: Age-standardised incidence and mortality rates for breast cancer in different world
regions ............................................................................................................................................. 52
Figure 3.1: Underlying principles of the interactive thresholding technique described by Byng et al.
A 10-bit digitiser (1024 greylevels) was used in the original study ......................................... 57
Figure 3.2: User interface for Cumulus v4 showing segmentation of breast from background (left)
and identification of dense tissue (right). ................................................................................... 57
Figure 3.3: Schematic example of visual analogue scale .......................................................... 58
Figure 3.4: Schematic diagrams of CC examination being performed (left) and resulting
mammogram (right). Arc A separates the outer breast edge from the background; Arc C
represents the last smooth iso-intensity curve. This represents the \( h_{12} = 0 \) curve ............... 62
Figure 3.5: Example calibration film from the original Manchester Method .............................. 64
Figure 3.6: Reference diagram showing co-ordinates \( x, y, z \) for Equation (3.4) ....................... 67
Figure 3.7: (a) SXA phantom; (b) the phantom as seen on the digital image ......................... 71
Figure 3.8: Digitised film priors (left) and processed digital mammography images (right) .... 72
Figure 3.9: The boundary between the compressed dense breast region and the layer of
subcutaneous fat in the breast edge is distinct and the skin edge appears as a bright white line. 73
Figure 3.10: Left to right: processed mammogram; Cumulus used by radiologist; Cumulus used by
physicist ............................................................................................................................................ 74
Figure 4.1: Geometry used to determine breast thickness, using radio-opaque markers placed on
the compression paddle ................................................................................................................ 78
Figure 4.2: Typical marker positions. The possibility of markers overlying breast tissue, or being obscured by other objects on the image must be minimised

Figure 4.3: Marker separation (in pixels) versus phantom thickness

Figure 4.4: Calibration data for marker pair 2 on 18x24cm field size. Pink: Bury site; Blue: Bolton site

Table 4.1: Marker calibration data, based on trend lines fitted to the graphs in Figure 4.3; \( m \) is the gradient and \( c \) is the intercept

Figure 4.5: Squash ball (left) and silicon-filled bra implant (right) compressed to 100N on the GE 800T unit to demonstrate paddle tilt

Figure 4.6: Variation in compressed breast thickness from the chest wall to the inner edge of the breast margin

Figure 4.7: Relationship between the difference in breast thickness over the compressed breast region and system-indicated breast thickness (left) and compression force (right)

Figure 4.8: Variation in compressed breast thickness due to deformation of the compression paddle in the lateral direction

Figure 4.9: Definition of stepwedge height, width and length

Figure 4.10: With no shielding, the grey level across each step varies due to x-rays passing through difference thicknesses of stepwedge material (left); lead shielding ensures that only those parts of the step where the x-rays have travelled through the full thickness are imaged (right)

Table 4.2: Properties of the materials considered for the new stepwedge compared to PTFE and breast tissue. All attenuation coefficient values are quoted at 20keV

Figure 4.11: Comparison of mass attenuation coefficients of aluminium, magnesium, breast tissue and PTFE over the energy range used in mammography

Table 4.3: Theoretical thickness of aluminium required to produce the same x-ray photon attenuation as a range of breast thicknesses at mean photon energy 20keV

Figure 4.12: (a) side view and (b) bird’s eye view of aluminium prototype showing the variation in step height and width and the location of notches to help identify each step

Figure 4.13: (a) Comparison of stepwedges. Top to bottom: 35mm PTFE, 25mm PTFE and aluminium prototype; (b) Stepwedges of 35mm PTFE (top) and aluminium (bottom) alongside 40mm Perspex

Figure 4.14: Optical density versus step height for aluminium and PTFE stepwedges

Figure 4.15: Aluminium prototype stepwedge exposed using 25kV, Mo/Mo, 135mAs

Figure 4.16: Optical densities on the stepwedge exposed alongside 2, 4 and 6cm Perspex. Exposure factors are shown in the legend

Figure 4.17: Characteristic curve for Kodak Min R-2000 film showing structure-oriented density values for a correctly exposed mammogram

Figure 4.18: Profiles of pixel value across the width of each step

Figure 4.19: ROI size and position

Table 4.4: Effect of step height and ROI area on signal (MPV) and noise (Std Dev)

Figure 4.20: Third aluminium prototype stepwedge: bird’s eye view (top) and side view (bottom)
Figure 4.21: Optical densities on the third prototype aluminium stepwedge exposed alongside 2, 4 and 6cm Perspex. Exposure factors are shown in the legend ................................................................. 97

Figure 4.22: (a) Stepwedge attached to breast support platform; (b) stepwedge remains securely in place during MLO projections; (c) stepwedge can be easily flipped out of the field of view ....... 98

Table 4.5: Comparison of density and linear attenuation coefficient for breast tissue .................. 99

Table 4.6: Comparison of density and linear attenuation coefficient for tissue equivalent material .......................................................................................................................................................... 100

Figure 4.23: Stepwedge thickness versus glandular thickness; legend shows total breast thickness

Data collected on (a) GE 800T at Bolton; (b) GE DMR at Bury ................................................ 101

Figure 4.24: Calibration surface generated in Matlab, The surface relates stepwedge thickness, glandular thickness and total breast thickness ........................................................................................................................... 102

Table 4.7: Typical glandularity variation with breast thickness for the screening population ...... 102

Figure 4.25: Optical density versus stepwedge thickness. The legend shows the total breast thickness with percentage glandularity in brackets .......................................................................................................................... 103

Figure 4.26: Breast image smoothed using a Gaussian filter. Blue: User-defined approximate breast outline; Yellow: horizontal profile, plotted in Figure 4.27 ................................................ 104

Figure 4.27: Resulting grey level profile for yellow line in Figure 4.26 ........................................ 105

Figure 4.28: Schematic illustration of radial profiles ............................................................................. 105

Figure 4.29: (a) original mammogram; (b) & (c) binary image before and after applying morphological operators .......................................................................................................................... 106

Figure 4.30: Side view of the breast in the CC position ................................................................. 107

Figure 4.31: (a) red dots indicate points on the breast edge and green lines show the normal profiles plotted to determine the breast margin; (b) smoothed breast edge; (c) results of thickness profiles drawn from the chest wall to the breast edge for four views from one examination ....... 108

Figure 4.32: Left: Original mammogram; Right: breast border defined by automatic segmentation .......................................................................................................................................................... 109

Figure 4.33: Left: Original mammogram; Right: user-defined pectoral muscle .................................. 109

Figure 4.34: Left: the user clicks on each of the markers; Right: a circle is fitted showing the centre of the marker .......................................................................................................................................................... 110

Figure 4.35: Hough transform for circle detection. The centre of a circle in the image domain is located by the peak signal in the parameter space ................................................................. 110

Figure 4.36: Examples where the Canny edge detector is unsuccessful and the user must manually select points on the edge of the marker .......................................................................................................................... 111

Figure 4.37: Compressed breast thickness profile. Red dots: actual data points based on the magnification markers; Green line: fit to data, extrapolated across image ................................................ 112

Figure 4.38: The user marks the edge of two steps and enters the numbers of these steps. ROIs in each step are shown so the user can check that they have correctly defined the steps .......... 112

Figure 4.39: Grey level values in each step .............................................................................................. 113

Figure 4.40: Left: original mammogram; Right: glandular thickness map ............................................ 114
Figure 4.41: Typical mammography spectrum of 28kV, Mo/Mo transmitted through 5mm aluminium (blue line) and 50mm soft tissue (pink line) .................................................................

Figure 5.1: Illustration of the anode heel effect (left) and the effect on optical density across the image, based on the experimental data of Cowen et al (right) ......................................................

Figure 5.2: (a) Original mammogram; Glandular thickness maps for (b) correct breast thickness, 33.2mm; (c) underestimated breast thickness, 24.3mm; (d) overestimated breast thickness, 43.4mm ............................................................................................................................................

Table 5.1: Effect of errors in breast thickness on percentage breast density ........................................ 120

Table 5.2: Effect of errors in breast thickness on the density estimate using Volpara™ ............... 121

Table 5.3: Coefficient of Variation (CV) for breast parameters generated by the analysis software .......................................................................................................................... 122

Figure 5.3: Image with the greatest intra-observer variation in breast area due to poorly defined pectoral muscle (left); possible operator selection of pectoral muscle (right) ......................... 123

Table 5.4: Images with the greatest observer variation in volumetric breast density ................. 124

Figure 5.4: Images with the greatest observer variation in volumetric breast density. 548RCC (left) and 504LML (right) .......................................................................................................................... 125

Figure 5.5: Bland-Altman plot for breast thickness ........................................................................ 126

Figure 5.6: Bland-Altman plot for breast volume ........................................................................ 127

Figure 5.7: Bland-Altman plot for glandular volume ..................................................................... 127

Figure 5.8: Bland-Altman plot for percentage breast density (%) .................................................. 128

Table 5.5: Inter-observer variation for breast parameters generated by the analysis software ... 129

Table 5.6: Images with the greatest discrepancy between observers. Shaded cells correspond to the blue arrows in Figures 5.5 – 5.8 ................................................................................................ 129

Table 5.7: Inter-observer variation for breast parameters generated by the analysis software ... 130

Figure 5.9: Image with the greatest observer variation in volumetric breast density, 806LML..... 131

Figure 6.1: Histograms of percentage mammographic density for two radiologists assessing a sample of 324 left breast images; the average of the LCC and LMLO densities are shown........ 134

Table 6.1: Visually assessed mammographic density ................................................................... 134

Table 6.2: Variation in mammographic density between radiologists ..................................... 134

Figure 6.2: Bland-Altman plot for visually-assessed mammographic density by two radiologists 135

Figure 6.3: Box and whisker plots of breast density, LCC view ..................................................... 137

Table 6.3: Correlation of mammographic density with volumetric breast density ......................... 138

Figure 6.4: Relationship between percentage volumetric breast density and mammographic density for the left breast. Linear and quadratic lines of fit are displayed.............................................. 139

Figure 6.5: Relationship between glandular volume and mammographic density for the left breast. Linear and quadratic lines of fit are displayed ................................................................. 139

Figure 6.6: Association between percentage breast density and the BI-RADS categories.
Classification is based on mammographic density assessment by Radiologist 2 ...................... 140

Table 6.4: Comparison of the distribution of percentage breast density data within the existing BI-RADS categories and the proposed categories based on linear and quadratic equations ..... 141
Figure 6.7: Original mammogram image and the six thresholds for classifying a dense pixel..... 143
Figure 6.8: Binary glandular tissue maps at each threshold for the image shown in Figure 6.7. Dense glandular regions are shown as white................................................................. 143
Table 6.5: Mean percentage dense area generated by the Manchester Method at each threshold and correlations with radiologist assessed mammographic density.............................................. 144
Figure 6.9: Bland-Altman plots for percentage dense area using Manchester Method with 10% threshold and radiologist visual assessment (left: Radiologist 1; right: Radiologist 2) ............... 145
Figure 6.10: Bland-Altman plots for percentage dense area using Manchester Method with 15% threshold and radiologist visual assessment (left: Radiologist 1; right: Radiologist 2) ............... 145
Figure 6.11: Bland-Altman plots for percentage dense area using Manchester Method with 20% threshold and radiologist visual assessment (left: Radiologist 1; right: Radiologist 2) ............... 145
Figure 7.1: (a) Stepwedge attached to breast support; (b) Stepwedge removed from field of view ................................................................................................................................................. 150
Table 7.1: Compressed breast thickness measured by the Manchester Method, by view .......... 152
Figure 7.2: Histogram of compressed breast thickness in the LCC view................................. 152
Table 7.2: Breast area, by view.................................................................................................................. 153
Figure 7.3: Histogram of breast area in the LCC view................................................................. 153
Table 7.3: Breast volume, by view........................................................................................................ 154
Figure 7.4: Histogram of breast volume in the LCC view........................................................... 154
Table 7.4: Glandular volume and percentage breast density, by view ........................................ 155
Figure 7.5: Histogram of glandular tissue volume in the LCC view........................................ 156
Figure 7.6: Histogram of percentage breast density in the LCC view........................................ 156
Figure 7.7: Box and whisker plots of compressed breast thickness by view ............................. 157
Table 7.5: Variation in compressed breast thickness, area and volume, by view ...................... 158
Figure 7.8: Schematic drawings of (a) Compressed breast in the CC view; (b) Same breast compressed in the MLO view............................................................................................................ 159
Figure 7.9: Box and whisker plots of breast volume by view ..................................................... 159
Figure 7.10: Relationship between breast volume and compressed breast thickness in the LMLO view............................................................................................................................................... 160
Figure 7.11: LMLO mammogram corresponding to the largest breast volume (3,398 cm³) .... 161
Table 7.6: Variation in glandular tissue volume and percentage breast density, by view .......... 161
Figure 7.12: Box and whisker plots of glandular tissue volume by view .................................... 162
Figure 7.13: Box and whisker plots of percentage breast density by view.................................. 163
Table 7.7: Breast thickness, volume and density for the cases shown in Figures 7.14 – 7.16 .... 163
Figure 7.14: Top: original mammogram and glandular thickness map for a breast with 46.7% density and glandular volume of 106.0cm³; Bottom: stepwedge pixel intensity plot..................... 164
Figure 7.15: Top: original mammogram and glandular thickness map for a breast with 57.9% density and glandular volume of 187.5cm³; Bottom: stepwedge pixel intensity plot..................... 165
Figure 7.16: Top: original mammogram and glandular thickness map for a breast with 29.0% density and glandular volume of 395.1cm³; Bottom: stepwedge pixel intensity plot..................... 166
Figure 7.17: Relationship between percentage breast density with compressed breast thickness (left) and breast volume (right) ................................................................. 167
Figure 7.18: Relationship between glandular tissue volume with compressed breast thickness (left) and breast volume (right) ................................................................. 168
Table 7.8: Number of missing images for each view ......................................................... 168
Figure 7.19: Pie-charts showing the proportion of cases for each method failure. The top chart for the RCC view includes the reason "sticker over stepwedge". If this reason is removed, the distribution of method failures is very similar for each view ............................................. 170
Figure 7.20: On the LMLO image (shown by red asterisk), the stepwedge obscures the breast. This is the only view affected within this examination and within the whole sample .......... 170
Table 7.9: Queries regarding the breast density study and responses given ......................... 174
Figure 7.21: Percentage breast density versus compressed breast thickness; the blue line is an estimation of the point at which the accuracy of breast density measurements becomes unreliable .................................................................................................................................. 176
Table 7.10: Pearson correlation coefficients between left and right breasts. SMF: Standard Mammographic Form; SXA: Single X-ray Absorptiometry ......................................................... 177
Table 8.1: Correlation of glandular volume with breast cancer risk factors ....................... 181
Table 8.2: Correlation of percentage breast density with breast cancer risk factors ........... 181
Figure 8.1: Relationship between glandular volume and age at examination .................. 183
Figure 8.2: Relationship between percentage breast density and age at examination ....... 183
Table 8.3: Correlation of percentage breast density and glandular volume with age, based on menopause status ................................................................. 184
Table 8.4: Hypothesis test summary for breast density distribution, based on menopause status .................................................................................................................. 184
Figure 8.3: Distribution of glandular volume in pre- and post-menopausal women ........... 185
Figure 8.4: Distribution of percentage breast density in pre- and post-menopausal women .... 185
Table 8.5: WHO classification of BMI ........................................................................... 186
Figure 8.5: Relationship between glandular volume and weight .................................... 186
Figure 8.6: Relationship between glandular volume and BMI category ......................... 187
Figure 8.7: Relationship between percentage breast density and weight ................. 187
Figure 8.8: Relationship between percentage breast density and BMI category .............. 188
Figure 8.9: Distribution of glandular volume in women with and without a family history of breast cancer ........................................................................................................... 189
Figure 8.10: Distribution of percentage breast density in women with and without a family history of breast cancer ........................................................................................................... 189
Figure 8.11: Relationship between glandular volume and years of menstruation .......... 191
Figure 8.12: Relationship between percentage breast density and years of menstruation .... 191
Figure 8.13: Distribution of glandular volume based on parity and age at first birth .......... 192
Figure 8.14: Distribution of percentage breast density based on parity and age at first birth... 193
Figure 8.15: Distribution of ethnicity within the sample .................................................. 194
Figure 8.16: Distribution of glandular volume based on lifetime HRT use .........................194
Figure 8.17: Distribution of percentage breast density based on lifetime HRT use .............195
Table 8.6: Descriptive statistics for years of HRT use and years since stopping ................195
Figure 8.18: Distribution of glandular volume based on current HRT use .........................196
Figure 8.19: Distribution of percentage breast density based on current HRT use .............196
Table 8.7: Hypothesis test summary for breast density distribution, based on current HRT use.197
Table 8.8: Use of Natural HRT remedies .............................................................................197
Table 8.9: Incidence of breast cancer within the study population ......................................198
Table 8.10: Variation in breast density between breast with cancer and contralateral breast .....198
Figure 8.20: Distribution of percentage breast density based on lifetime breast cancer diagnosis ..........................................................................................................................................................................................199
Figure 8.21: Histograms of LSOA rank for (a) Questionnaire 2 respondents and (b) Bolton and Bury ..........................................................................................................................................................................................................................................................201
Figure 8.22: Relationship between percentage breast density and LSOA rank ......................201
Table 8.11: Definitions of urban / rural indicators and the distribution of the sample within each 202
Figure 8.23: Histogram of radiologist-assessed percentage mammographic density ..........203
Table 8.12: Correlation of percentage breast density (left breast) by area and by volume with breast cancer risk factors ..........................................................................................................................................................................................................................................................203
Table 8.13: Comparison of the distribution and mean BMI between published population data and the study population data ..........................................................................................................................................................................................................................................................204
Table 8.14: General demographics and results of paired t-tests for women who completed both questionnaires ..........................................................................................................................................................................................................................................................205
Table 8.15: Summary of responses from Questionnaire 1 and associated level of confidence...206
Figure 8.24: Relationship between mean weight and level of certainty in response ............207
Table 8.16: Correlation of breast density in the CC view with BMI and weight ....................208
Table 8.17: Correlation of breast density in the CC view with parity ....................................209
Figure 9.1: Screen-shot of the Fuji Profect CR system, illustrating the extensive range of image processing algorithms available for QC. There are additional options for clinical images........213
Figure 9.2: Detector response for (a) GE Essential with a-Si detector; (b) Hologic Selenia Dimensions with a-Se detector ..........................................................................................................................215
Figure 9.3: Relationship between glandular thickness and ln(MPV/mAs) on the Hologic Selenia; the legend shows total breast thickness and kV/T/F and the trendline shows the relationship between these variables and the strength of fit ..........................................................................................................................................................................................................................................................216
Figure 9.4: Relationship between glandular thickness and ln(MPV/mAs) on the GE Essential for a range of kV and T/F combinations; the legend shows total breast thickness and the trendline shows the relationship between these variables and the strength of fit .................................................................217
Table 9.1: Gradient and intercept of the linear fits to the data for 40mm phantom thickness, demonstrating the effect of kV and T/F ..........................................................................................................................................................................................................................................................218
Table 9.2: Factors selected under automatic exposure control which could be used to inform the range of kV and T/F permutations required for calibration data ..........................................................................................................................................................................................................................................................218
Figure 9.5: DICOM header for GE Essential. Relevant DICOM tags are highlighted..................219
Figure 9.6: Comparison of the detector response for 6 GE Essential units (left) and 6 Hologic Selenia Dimensions units (right) ..........................................................................................................................220
Figure 10.1: Example of overestimating glandular thickness within the breast margin............227
Figure 10.2: Digital mammography images with poor tissue thickness equalisation; the breast is clearly separated into the compressed breast region and the margin.........................................................228
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACRIN</td>
<td>American College of Radiology Imaging Network</td>
</tr>
<tr>
<td>AEC</td>
<td>Automatic Exposure Control</td>
</tr>
<tr>
<td>BI-RADS</td>
<td>Breast Imaging Reporting and Data System</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>CC</td>
<td>Cranio-caudal</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CR</td>
<td>Computed Radiography</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of Variation</td>
</tr>
<tr>
<td>DBCT</td>
<td>Dedicated Breast Computed Tomography</td>
</tr>
<tr>
<td>DCIS</td>
<td>Ductal Carcinoma In Situ</td>
</tr>
<tr>
<td>DICOM</td>
<td>Digital Imaging and Communications in Medicine</td>
</tr>
<tr>
<td>DMIST</td>
<td>Digital Mammographic Imaging Screening Trial</td>
</tr>
<tr>
<td>DR</td>
<td>Digital Radiography</td>
</tr>
<tr>
<td>DIXA</td>
<td>Dual X-ray Absorptiometry</td>
</tr>
<tr>
<td>FFDM</td>
<td>Full Field Digital Mammography</td>
</tr>
<tr>
<td>FFTP</td>
<td>First Full Term Pregnancy</td>
</tr>
<tr>
<td>HRT</td>
<td>Hormone Replacement Therapy</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass Correlation Coefficient</td>
</tr>
<tr>
<td>ICRU</td>
<td>International Commission on Radiological Units and Measurements</td>
</tr>
<tr>
<td>IMD</td>
<td>Indices of Multiple Deprivation</td>
</tr>
<tr>
<td>IPEM</td>
<td>Institute of Physics and Engineering in Medicine</td>
</tr>
<tr>
<td>LSOA</td>
<td>Lower-layer Super Output Area</td>
</tr>
<tr>
<td>MD</td>
<td>Mammographic Density</td>
</tr>
<tr>
<td>MLO</td>
<td>Medio-lateral oblique</td>
</tr>
<tr>
<td>MPV</td>
<td>Mean Pixel Value</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>NHSBSP</td>
<td>National Health Service Breast Screening Programme</td>
</tr>
<tr>
<td>OD</td>
<td>Optical Density</td>
</tr>
<tr>
<td>OR</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>PACS</td>
<td>Picture Archiving and Communication System</td>
</tr>
<tr>
<td>PMD</td>
<td>Percentage Mammographic Density</td>
</tr>
<tr>
<td>PROCAS</td>
<td>Predicting Risk of Cancer At Screening</td>
</tr>
<tr>
<td>PTFE</td>
<td>Polytetrafluoroethylene (Teflon)</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>SCC</td>
<td>Six Category Classification (Boyd)</td>
</tr>
<tr>
<td>SES</td>
<td>Socioeconomic Status</td>
</tr>
<tr>
<td>SMF</td>
<td>Standard Mammographic Form</td>
</tr>
<tr>
<td>SXA</td>
<td>Single X-ray Absorptiometry</td>
</tr>
<tr>
<td>TDLU</td>
<td>Terminal Duct Lobular Unit</td>
</tr>
<tr>
<td>T / F</td>
<td>Target / Filter</td>
</tr>
</tbody>
</table>
Abstract

The University of Manchester
Jennifer Louise Diffey
Doctor of Philosophy (PhD)
Development and evaluation of a method for measuring breast density
24 September 2012

Introduction: Breast density is an important independent risk factor for breast cancer and is negatively associated with diagnostic sensitivity of mammography. Measurement of breast density can be used to identify women at increased risk of developing breast cancer and those who would benefit from additional imaging. However, measurement techniques are generally subjective and do not reflect the true three-dimensional nature of the breast and its component tissues.

Method: A semi-automated method for determining the volume of glandular tissue from digitised mammograms has been developed in Manchester. It requires a calibration device (stepwedge) to be imaged alongside the breast during mammography, with magnification markers on the compression paddle to accurately determine breast thickness. Improvements to the design of the stepwedge and markers have enabled the method to be applied to the screening population for the first time. 1,289 women had their volumetric breast density measured using this method and additionally completed a questionnaire on breast cancer risk factors.

Results: The method has demonstrated excellent intra- and inter-observer agreement. The median percentage breast density in the study cohort was 8.4% (interquartile range 4.9 – 14.2%). There was no significant difference between left and right breasts; the difference between MLO and CC views was significant (CC view was denser), but values were closely correlated (r = 0.92, p < 0.001). The median glandular volume was 60.1cm$^3$ and exhibited no significant variation between left/right breasts or CC/MLO views. A number of breast cancer risk factors were found to be significantly correlated with glandular volume and percentage breast density, including age, weight, BMI, parity, current HRT use and current smoking. The strength of correlation was equal to or greater than that of visually assessed mammographic density. Glandular volume and percentage breast density measurements demonstrated strong relationships with visually assessed mammographic density, which has been shown to be highly correlated with risk.

Conclusions: These findings are promising and suggest that volumetric breast density measured using this method should be associated with breast cancer risk. However, further work is required to establish this relationship directly. The method will be used in a large study, known as PROCAS (Predicting Risk Of Cancer At Screening) which aims to develop individualised breast cancer risk prediction models; these have the potential to form the basis of tailored screening intervals. Preliminary work has been undertaken to adapt the method for full field digital mammography, which suggests that it is possible to use the integrated digital detector as the calibration device.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

During the course of this PhD, five medical students carried out 11 week projects under my co-supervision. Some of the data collected for their studies have been included in this thesis but all analysis and interpretation is my own. The medical students and their contributions are listed below, and their contributions are acknowledged at the appropriate points in the thesis.

Joanne Greene (2008) assisted with entering data from Questionnaire 1 and arranged for Dr Tony Maxwell to read mammographic density on the mammograms of 196 women using the visual analogue scale. Joanne converted the markings on the scale to numerical values.

Camilla Jeffries-Chung (2009) arranged for Dr Mary Wilson to visually assess mammographic density on the mammograms of 273 women using the visual analogue scale. Camilla converted the markings on the scale to numerical values.

Joanna Morrison (2009) ran the analysis software on all digitised cranio-caudal mammograms collected during the feasibility study. This generated results of compressed breast thickness, breast area, breast volume, glandular tissue volume and percentage breast density, which are included in the thesis.

Rosanne Verow (2009) carried out an inter-observer analysis study. As part of this, Rosanne ran the analysis software on a number of medio-lateral oblique images. These are the only results that have been used in the inter-observer study presented in the thesis.

Rita Prajapati (2010) compared the responses and associated confidence levels between the original questionnaire issued to the participants of the feasibility study in 2007 and a modified questionnaire issued in 2009.

Additionally, Michael Berks (a research associate at ISBE) is credited with the development of the algorithm used to detect the breast margin and with assistance in programming using Matlab (The Mathworks, Inc).
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
Acknowledgements

I would like to say a very special thank you to my supervisors, Dr Sue Astley and Dr Alan Hufton, for their guidance and encouragement throughout the PhD and their valuable feedback on the thesis. It would not have been possible without their support.

John Lewis plc and Genesis are gratefully acknowledged for providing the funding for the feasibility study. I am extremely appreciative to all staff at Bolton, Bury and Rochdale Breast Screening Service for their huge assistance with the feasibility study. A particular mention must go to Claire Mercer (Superintendent Radiographer), Dr Tony Maxwell (Clinical Director / Chief Radiologist), Susan Butler (Office Manager) and Jackie Kirby (Bury Receptionist). Thanks also go to Tom Hamnett at the Christie Hospital, Manchester for constructing the stepwedges and marker sheets required for the study.

I would like to express thanks to everyone who assisted me with the data entry and digitisation of mammograms associated with the feasibility study, especially Sandra Danquah, Joanne Greene, Joanna Morrison, Camilla Jeffries-Chung, Rosanne Verow and Alex Fitton. Many thanks also go to my colleague, Michael Berks, for his patient and knowledgeable assistance with the Matlab programming.

I am grateful to Dr Tony Maxwell and Dr Mary Wilson for giving up their time to assess mammographic density on a number of mammograms from the feasibility study.

Finally, I would like to thank my family and my boyfriend Garth, for having confidence in me and cheering me on in the final months.
The Author

Jennifer Diffey graduated from the University of Leeds with a BSc (Hons) in Physics (2000) and an MSc in Medical Physics (2003). Her career to date has been in the field of medical physics. She completed her training at the Leeds Teaching Hospitals and the Christie Hospital in Manchester, where she specialised in Diagnostic Radiology and Radiation Protection. She developed a particular interest in mammography and took on the role of the North West representative for the NHSBSP Co-ordinating Group for the Physics of Mammography. In 2009, she moved to Sydney where she acts as the State Medical Physicist for BreastScreen NSW.

Her research experience consists primarily of the work completed for this PhD, although she has collaborated with other physicists and radiographers and anticipates that these opportunities will continue.

Her publications are listed below. Those which are not directly related to this PhD are in italics.


1. Introduction

Breast cancer is a major health burden on a global scale. In 2008 there were 47,693 new cases diagnosed in women in the UK, and 341 in men. Breast cancer accounts for 31% of all cases in women, making it by far the commonest cancer in women and perhaps surprisingly, the most common cancer in the UK [1]. Since 2000, over 1 million new cases of female breast cancer have been diagnosed worldwide each year. The figure reached 1.38 million in 2008 [2] and is now estimated at close to 1.5 million.

Since the introduction of the National Health Service Breast Screening Programme (NHSBSP) in 1988, mortality rates have decreased steadily. Nevertheless, breast cancer still accounts for 16% of female mortality from cancer in the UK, with 12,047 deaths in 2008 [1]. However, screening by x-ray mammography is not infallible, having a sensitivity of typically 85% [3]. This means that 15% of cancers are missed overall (false negatives), with a greater percentage being missed in women with high mammographic density. Screening has been criticised in recent years for overdiagnosis and overtreatment of disease, resulting in unnecessary psychological stress and an increase in the number of invasive and surgical procedures, including mastectomies [4, 5]. Some authors claim that at least two-thirds of prevented deaths are explained solely by advances in treatment [6, 7]. Inevitably, these claims are heavily contested and recent large scale studies have shown a 31% reduction in breast cancer mortality due to screening, although the authors make it clear that it is necessary to have follow-up times of at least 20 years in order to fully observe such benefits [8, 9].

For UK women, the lifetime risk of developing breast cancer is currently estimated as 1 in 8 [10]. However, this can vary greatly between individuals, depending on a number of risk factors, such as age, weight and hormonal factors. In recent years, there has been increasing interest in breast density as a risk factor, which is the focus of this PhD. It is possible that the efficacy of routine breast screening could be improved by tailoring the screening interval to individual risk [11]. Currently, age is the only criterion for entry into the screening programme, and screening intervals are fixed at three years for all women, except for those with a family history of breast cancer who are screened annually.

1.1 Breast density

In 1976 Wolfe [12, 13] proposed that the composition of the breast, based on the parenchyma, was related to the risk of breast cancer. His study was prompted by the observation that with increasing age, the connective and epithelial tissues were seen to regress, causing the prominent ducts to become the dominant feature. He hypothesised that the prominent duct pattern was associated with the presence of breast carcinoma, a history of breast carcinoma and an increased
risk of breast cancer. Although this was true to a certain extent, it was subsequently recognised that mammary dysplasia was found to have an even stronger association with breast cancer risk. Wolfe used the term mammary dysplasia to indicate tissue with a radiological density greater than fat.

Most studies investigating the association of breast density with breast cancer risk regard the breast as a simple two-compartment model, defining tissue as either fat or fibroglandular. Fat attenuates x-rays to a lesser extent than fibroglandular tissue and appears dark on the mammogram, whereas the dysplasia, or glandular tissue, is radiodense and appears light. The area of the mammogram occupied by the radiologically dense breast tissue is now commonly referred to as mammographic density (MD). This can be expressed as a proportion of the total breast area to give the percentage mammographic density (PMD) and it is this variable which has been employed in most clinical studies to date. These concepts are demonstrated in Figure 1.1 which shows examples of mammograms within each group of the Boyd 'six category classification' (SCC) scheme [14, 15].

![Figure 1.1: Boyd categories for mammographic density](Reproduced with permission from [15], Copyright Massachusetts Medical Society)

Mammographic density is a strong risk factor for breast cancer. Early studies pioneered by Wolfe and Boyd [12-14, 16] inspired further research and over the past three decades, more than 50 studies have been published which investigate the relationship between mammographic density...
and breast cancer risk [17]. Inevitably, the strength of the association varies between studies, but a four-fold increase in risk appears typical for women in the highest, compared to the lowest category of mammographic density [18]. The magnitude of the risk related to mammographic density is greater than most other risk factors, leading one author to describe it as being “perhaps the most undervalued and under-utilised risk factor in studies investigating the causes of breast cancer” (Byrne, 1997) [19].

Women with breast carcinoma are often found to exhibit highly dense breasts and it can be very difficult to diagnose the disease in these women, leading some to propose that the apparent association between breast cancer risk and breast density is in fact an artefact, due to masking bias [20]. Cancers missed at the first screen because they were obscured by radiologically dense tissue would be detected shortly thereafter as an interval cancer. If a study included only these women as the ‘cases’, this would give the erroneous impression that incidence is higher in women with mammographically dense breasts [17, 19]. A number of large-scale studies [18, 21, 22] have shown that although masking bias is a real effect, it can be accounted for by long follow up periods and the inclusion of subsequent screen-detected cancers; the strong association between breast cancer incidence and extensive mammographic density still remains.

The reasons for increased breast cancer risk associated with extensive mammographic density are not fully understood. Twin studies [15] have shown that heritability accounts for 63% of the variation in mammographic density observed in twins, but no ‘breast density gene’ has been identified. A number of studies have examined the relationship between mammographic density and other risk factors, in order to better understand why some women have dense breasts and why they are at increased risk of breast cancer. It is interesting to note that despite being strongly correlated with a number of other risk factors, after accounting for these confounding factors in clinical studies, association of mammographic density with breast cancer risk remains strong, making it an independent risk factor for the disease. Consequently, there is interest in including mammographic density in individualised risk prediction models. Well-established models include the Gail [23], Claus [24] and Ford [25, 26] models but these only have a limited number of input variables. Barlow et al [27] found that mammographic density, classified using the BI-RADS categories [28], was a statistically significant risk factor for breast cancer diagnosis in pre- and post-menopausal women and its inclusion in risk prediction models may offer improved accuracy in identifying women at high risk of developing cancer.

It is important to note that the vast majority of studies published to date have used mammographic density (projected dense area), or more commonly, percentage mammographic density as the risk factor and have generally treated PMD as a categorical variable rather than a continuous variable. Because of the number of confounding factors associated with mammographic density, it has been suggested that square-root dense area is in fact the best single predictor [29]. Even if this were the case, the scientific validity of using any area-based parameter has been questioned [30,
Although the density versus risk mechanism is not fully known, it is logical that risk is more closely associated with the actual amount of dense tissue within the breast rather than its projected area [32]. Mammographic density is essentially based on film optical densities from a two-dimensional projection image of the breast and is not a true measure of the tissue volumes that created those densities. Such techniques are subjective and suffer from errors associated with inconsistencies in breast positioning. Additionally, they cannot account for depth information or variations in x-ray exposure.

There is a clear need for volumetric techniques which truly measure breast density. A number of methods exist, including the one employed in this thesis, which is a calibration technique known as the Manchester Method (Chapter 4). The Manchester Method requires a stepwedge to be imaged alongside the breast during every mammogram. The grey level of each pixel in the breast image is matched to the equivalent grey level in the stepwedge. Using previously generated calibration data, the corresponding thickness of stepwedge material is used to determine the thicknesses of glandular and adipose tissue within the column of tissue projected onto that pixel.

Several volumetric methods have been validated using tissue-equivalent phantoms, or by examining the correlation with area-based techniques that have shown a strong and well-established link to breast cancer risk, but clinical evidence on the relationship between volumetric breast density and risk is sparse. Two small case-control studies using film-screen mammograms have been reported using a model-based [33] and a calibration technique [34] to estimate volumetric breast density. In both studies, the association between percentage volumetric density and breast cancer risk was found to be lower than the corresponding area-based measure. When both measures were included in a predictive regression model, only the area-based measure retained significance. This is somewhat disappointing, given that the volume of dense tissue is hypothesised to be more closely related to breast cancer risk. However, some volumetric techniques are more likely to fail when applied to very dense breasts and excluding these from the sample will inevitably result in an underestimation of risk. This is certainly an issue that needs to be resolved.

One of the most interesting aspects of breast density is that it can be reduced by interventional methods, such as diet, exercise and the use of oestrogen receptor modulators. However, it is not yet known if a reduction in density carries a corresponding reduction in risk. Until this has been firmly established, breast density remains a potential, but unconfirmed, surrogate marker for breast cancer risk [35]. In order to assess the effect of intervention on density and risk, it is necessary to develop techniques which are objective, reproducible and capable of accurately measuring small changes in density over time. It is thought that only volumetric techniques will fulfil these criteria.
1.2 Aims and objectives

The aims of this PhD are expressed in the headings below, accompanied by a short explanatory paragraph and details of where these aims are addressed in the thesis.

1.2.1 Extend the Manchester Method for volumetric breast density measurement

Many principles of the Manchester Method have already been established [36 - 38]. However, the original method suffered from a number of limitations which meant that it had previously only been applied to 39 women taking part in a 'lifestyle study', examining the effect of diet and exercise on breast density and volume. Significant improvements to the stepwedge are required in order to facilitate the use of the method in the breast screening programme.

The design of a new, compact stepwedge is the main aim of method extension. Additionally, improvements to the estimation of breast thickness are required, both in the compressed breast region where this can be accurately measured, and in the breast margin where thickness must be modelled. Chapter 4 describes the advancement of the method and provides a full discussion of how these improvements were achieved.

1.2.2 Validate the new method

The method must be validated prior to large scale clinical use. Low intra- and inter-observer variability is essential to guarantee reliability of the technique. Furthermore, uncertainties associated with each of the steps involved in the calibration process must be quantified and steps taken to ensure that these are minimised. This is presented in Chapter 5.

In general terms, validation is defined as the process taken to ensure that a technique meets the requirements for which it was developed. In specific terms, this means that the volumetric breast density measurements generated by the Manchester Method must exhibit a strong association with breast cancer risk, or with known risk factors. The feasibility study described below (Aim 1.2.5) will therefore contribute to the validation process. An additional strategy to determine the suitability of the Manchester Method as a risk assessment tool is to examine the correlation of breast density measurements with those from a technique which has shown a strong and well-established association with breast cancer risk. This is the approach described in Chapter 6, where results from the Manchester Method are compared to visually assessed mammographic density.
1.2.3 Evaluate the feasibility of using the improved method in the breast screening programme

The original method had only been applied to a very small number of participants in a research environment. However, in order to fully realise the potential of volumetric breast density measurement, it would be preferable to apply the technique in the screening programme. Any method employed must not disrupt the clinical examination and radiographer input must be kept to an absolute minimum. Furthermore, the method must be capable of providing a measure of breast density for all breast sizes and compositions. Details of the feasibility study are given in Chapter 7.

1.2.4 Obtain information on the distribution of volumetric breast density in the screening population

The women taking part in the 'lifestyle study' were not necessarily representative of the screening population as they were generally overweight, younger and pre-menopausal.

The application of the Manchester Method to the screening population will enable the collection of sufficient data to establish relationships between breast thickness, breast volume, glandular tissue volume and percentage breast density. It will be interesting to examine the range of values generated by the Manchester Method and to compare the findings with results from other published methods. Descriptive statistics on breast density are presented in Chapter 7.

1.2.5 Correlate absolute glandular volume and percentage breast density with established breast cancer risk factors and risk

An additional aim of the feasibility study is to correlate breast density (absolute and percentage glandular volume) with breast cancer risk factors such as age, weight, family history, parity and use of HRT. This had not previously been done using the Manchester Method and literature on the association of risk factors with volumetric breast density from other published methods is also limited. Women invited to take part in the feasibility study would be provided with a questionnaire requesting this risk-related information. It is also anticipated that sufficient data will be collected to examine the association between breast density measured using the Manchester Method and breast cancer risk. The results of these analyses are provided in Chapter 8.

1.2.6 Adapt the method to make it suitable for use in digital mammography

Full field digital mammography (FFDM) is rapidly replacing film-screen mammography in the UK screening programme and breast density techniques must be adapted to reflect the changing technology. Integrated digital mammography detectors exhibit a linear response between grey level value (on the raw image) and detector dose. This relationship is advantageous for volumetric
methods, particularly calibration techniques such as the Manchester Method where it may be possible to use the detector itself as the calibration device, rather than requiring a separate phantom in the image.

Preliminary work on adapting the Manchester Method for digital mammography is included in Chapter 9.

1.2.7 Adapt the method to measure density by area and examine the association with visually assessed density

Mammographic density is not a true reflection of breast density and a pixel which is classed as dense will not be composed entirely of glandular tissue. It would be useful to gain an insight into the true tissue composition of such a pixel as this may enable the derivation of a simple relationship between mammographic density and volumetric breast density.

The Manchester Method provides an estimate of the total breast and glandular tissue thicknesses for every pixel in the digitised image. The thickness of glandular tissue could be expressed as a percentage of the total breast thickness and a threshold applied to classify this pixel as ‘dense’. The number of dense pixels would be summed and expressed as a percentage of the number of pixels within the breast, thereby providing an area-based measure which can be compared to mammographic density. The aim of doing this is to simulate the visual assessment process and determine which threshold corresponds most closely to the way in which radiologists classify dense pixels. This study is described in Chapter 6.

1.3 Thesis outline

Detailed background information to the study is provided in Chapter 2. Breast density classification schemes are introduced as well as their application to large-scale clinical studies, from which the magnitude of the risk associated with breast density has been derived. A number of breast cancer risk factors are presented and their relationship with mammographic density is described, where applicable. Note that their relationship with volumetric breast density requires further investigation. Chapter 3 provides a critical analysis of quantitative methods for breast density measurement, both by area and by volume. The challenges of adapting these techniques for digital mammography are also discussed.

Chapters 4 – 9 contain the original experimental work carried out for this PhD. Chapter 4 describes the evolution of the Manchester Method, with particular emphasis on the design of a new stepwedge and refinement of the marker sheets. The method is semi-automated and the description of each stage of the method is accompanied by screenshots from the analysis
software. This chapter also describes studies undertaken to evaluate the new stepwedge and to assess the extent of paddle tilt.

In Chapter 5, an analysis is presented on the effect of uncertainties in the Manchester Method on the accuracy of volumetric breast density measurements. An assessment of the reliability of the method is also described, based on intra- and inter-observer variability studies. The validation processes are discussed in Chapter 6 and the measurements of breast density using the Manchester Method (by volume and area) are compared to those of visually assessed mammographic density.

The feasibility study is presented in Chapters 7 and 8. Chapter 7 includes descriptive statistics on breast thickness, area and volume as well as glandular tissue volume and percentage breast density; the relationships within and between these measures are also evaluated. Method failures are examined in detail and suggestions are made for how these could be prevented in future studies.

Correlations between volumetric breast density and breast cancer risk factors are shown in Chapter 8. Both uni-variable and multivariate analysis has been carried out. The association between breast density and breast cancer risk is also investigated. All risk factor data were self-reported and the validity of this approach is discussed.

Chapter 9 describes preliminary work on the adaptation of the Manchester Method to digital mammography, including a study on the stability of digital detectors.

The conclusions from this study and the contributions to learning are presented in Chapter 10. Areas of future work are also proposed.
2. Background

This chapter introduces a number of topics related to the work in this thesis, starting with a description of the UK screening programme and the underlying physical principles of mammography. Details of breast anatomy and breast parenchyma are presented. This is followed by a comprehensive discussion on classification schemes for breast density and how their application in clinical studies has established breast density as a strong risk factor for breast cancer. Finally, information is presented on breast cancer risk factors, their relationship with breast density and their use in risk prediction models.

2.1 National Health Service Breast Screening Programme (NHSBSP)

Figure 2.1 shows the female breast cancer incidence and mortality trends for the UK from 1975 to 2008. In 1988, the world’s first national breast screening programme was set up in England, following the recommendations of the 1986 Forrest Report [39]. It is reassuring to observe that mortality rates have decreased steadily since the introduction of the NHSBSP, although advances in treatment must also be acknowledged. As the population ages, risk and consequently incidence are expected to rise, making breast screening as important now as when it was introduced. Note that the sharp increase in incidence in Figure 2.1 results from the screen-detection of a prevalent pool of undiagnosed cancers.

![Figure 2.1: Breast Cancer, European Age-Standardised Incidence and Mortality Rates, UK Females, 1975-2008 [1]](image)
Initially, women aged 50 – 64 were invited for screening every three years. At their first (prevalent) screen, they received two-view mammography (cranio-caudal and medio-lateral oblique views of each breast) and at subsequent (incident) screens, they received medio-lateral oblique views only. In 2004, the invited age range was extended to 69 years, with women over the age of 70 eligible for self-referral. The screening interval was unchanged but two-view mammography was used at every screen, having been phased in over the previous two years. By 2016, it is anticipated that the invited age range will be extended further to 47 – 73 [10]. Some important statistics on the NHSBSP, taken from the 2010-11 report [40] are shown in Table 2.1.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>2010-11</th>
<th>2009-10</th>
<th>2000-01</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number screened</td>
<td>1.88 million (2.44 million invitations sent)</td>
<td>1.79 million</td>
<td>1.30 million</td>
<td>Increase due to age extensions of programme and population growth in screening age range</td>
</tr>
<tr>
<td>Uptake</td>
<td>73.4</td>
<td>73.2</td>
<td>Must be ≥ 70% [41]. Greater uptake amongst previous attendees (87.5%) than first-time invitees (69%)</td>
<td></td>
</tr>
<tr>
<td>No. of cancers detected</td>
<td>14,725</td>
<td>14,229</td>
<td>8,345</td>
<td>Increase due to age extensions of programme, population growth in screening age range and two-view screening</td>
</tr>
<tr>
<td>Detection rate per 1,000</td>
<td>7.8</td>
<td>7.9</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>Cancer detection rate per 1000, by age (2010-11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 - 64</td>
<td>6.2</td>
<td>8.4</td>
<td>10.1</td>
<td>14.7</td>
</tr>
<tr>
<td>Invasive (all)</td>
<td>Invasive (&lt; 15mm)</td>
<td>Non/micro-invasive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer type, % (2010-11)</td>
<td>79.5</td>
<td>40.4</td>
<td>20.5</td>
<td>&lt; 15mm diameter is too small to detect by self-examination</td>
</tr>
</tbody>
</table>

Table 2.1: Programme statistics (England)
Compiled from the NHSBSP Statistical Bulletin, 2010-11 [40]

It has been estimated that if 70% of eligible women attend screening, which is the minimum standard set by the NHSBSP [41], there would be a 25% reduction in breast cancer mortality rates in those women invited for screening. In England, the breast screening programme currently costs around £96 million a year [10].

The sensitivity of mammography in women aged over 50 has been found to range from 68% to over 90%, depending on the screening interval and the time elapsed since mammography. Most
programmes achieve sensitivities of around 85%, which means that 15% of cancers are missed (false negatives). In women aged 40-49 the sensitivity is lower, with estimates between 62% and 76%. This is because pre-menopausal women have more dense breast tissue which is likely to obscure areas of abnormality on a mammogram. The specificity of breast screening by mammography is between 82% and 97% [3].

In recent years, breast screening has been a source of controversy in the literature. Some studies estimate that one life would be saved for every 300 women screened over 10 years, resulting in a 31% reduction in breast cancer mortality due to screening [8, 9], whilst others claim that two-thirds [6] to all [7] deaths prevented are explained by advances in treatment. Overdiagnosis is acknowledged as a product of screening. This refers to the detection of disease that, ultimately, will not cause symptoms or early death. However, it is thought that 2 – 2.5 lives are saved for every case estimated to have been overdiagnosed [8, 42]. This is considered unacceptable by some, such as the Nordic Cochrane Review who state that “for every 2,000 women invited throughout 10 years, 1 will have her life prolonged, 10 will be treated unnecessarily and over 200 will experience psychological stress for months due to false positive findings [5].” A further criticism is overtreatment of the disease, with arguments that certain screen-detected cancers, especially ductal carcinoma in situ (DCIS), would never become clinically evident and may even spontaneously regress. However, this is thought to be rare and it is therefore unethical not to provide follow-up examination and treatment to these women [42].

Interpreting the literature is difficult as differences in methodology, particularly follow-up time, can often explain the contradictory results [43]. Scientific studies are often misreported by journalists and women are increasingly confused about the benefits of breast screening [44]. As a result, an independent review of the NHSBSP commenced in October 2011, led by Professor Sir Mike Richards, the National Cancer Director [45].

It has been proposed that the effectiveness of routine breast screening could be improved by tailoring the screening interval to individual risk. This approach was recently modelled by a US group led by Schousboe et al [11], who compared annual, biennial and 3 – 4 yearly mammography in the age groups 40 – 49, 50 – 59, 60 – 69 and 70 – 79. A number of risk factors were taken into account and the outcome measures were the costs per quality-adjusted life-year and the number of women screened over 10 years to prevent one death from breast cancer. They concluded that women should receive their first mammogram at age 40 in order to categorise their breast density. Future screening intervals should be personalised on the basis of age, breast density, history of breast biopsy and family history of breast cancer.
2.2 Physical principles of mammography

Mammography is the x-ray examination of the breast, shown schematically in Figure 2.2 (left). It is carried out using dedicated equipment, such as the digital mammography unit shown in Figure 2.2 (right). The ‘image receptor’ in this case is an integrated amorphous selenium (a-Se) flat panel detector. Analogue units would have a slot at this location for inserting the cassette. The cassette could contain film-screen or a Computed Radiography (CR) phosphor plate. Film-screen systems are gradually being replaced with Full Field Digital Mammography equipment (FFDM) in the UK screening programme. FFDM encompasses a range of detector technologies, which are discussed in detail in Chapter 9 (Adaptation of the Manchester Method to FFDM). Examples include Computed Radiography (CR), which is used with existing analogue x-ray equipment, and integrated digital solutions (DR) which utilise flat panel technology or photon-counting detectors.

![Diagram of mammography equipment](image)

*Figure 2.2: Left: Schematic representation of mammography examination; Right: typical mammography x-ray unit [46]*

The demands on mammography are stringent: both high spatial resolution and high contrast are required to detect microcalcifications and to visualise masses respectively. Radiation doses must be kept as low as reasonably achievable (ALARA), a requirement which is absolutely paramount for screening where approximately 99% of the population will be cancer-free at the time of their examination.

Malignant tissue has negligible physical density difference from fibrous breast tissue and there is only a small difference between the photon interaction properties of the two, resulting in very low subject contrast. Detection of carcinoma is therefore difficult. Figure 2.3 shows the subject...
contrast of malignant breast tissue and calcification, relative to fibrous tissue. The values have been calculated using the experimental data of Johns and Yaffe [47], who measured attenuation coefficients and densities of fat, fibrous tissue and infiltrating duct carcinoma.

![Graph showing subject contrast of malignant tissue and calcification, relative to fibrous tissue.](image)

*Figure 2.3: Subject contrast of malignant tissue and calcification, relative to fibrous tissue. Based on experimental data from [47]*

However, malignant tissue does have a higher effective atomic number (Z) than fibrous tissue. Mammography therefore utilises the photoelectric effect, which is proportional to $Z^3 / E^3$, to enhance the contrast between these tissues, based on this difference in atomic number. Low x-ray beam energy, $E$, which is well-suited to soft-tissue imaging, also increases the likelihood of photoelectric interaction [48, 49]. Maximum beam energy will depend upon breast thickness and glandularity but is typically 25 – 35keV.

The most common x-ray spectrum used for film-screen mammography is generated using a molybdenum (Mo) target with a Mo filter, as shown in Figure 2.4 [50]. A considerable portion of the emission spectrum is due to K-characteristic radiation at 17.4 and 19.6keV. The filter not only removes the unnecessary low energy bremsstrahlung (continuous) radiation, which would contribute to breast dose but not to the image, but also significantly reduces the intensity of the higher energy radiation, improving contrast. There is a sharp reduction at 20.0keV which is the K-absorption edge of molybdenum. As breast thickness and density increases, a more penetrating x-ray beam is required. This is achieved by using a rhodium (Rh) filter, which removes a greater proportion of the low-energy bremsstrahlung radiation and greatly attenuates radiation at the K-absorption edge of 23.2keV. Some mammography units also have a Rh target, used exclusively with a Rh filter. The Rh/Rh spectrum has K-characteristic x-rays at 20.2 and 22.8keV and is only selected for breasts which are particularly thick and / or dense [48, 49].
Because the mammogram is a two-dimensional projection image of the breast, it is essential to compress the breast. Compression serves a number of functions pertaining to increased image quality and lower radiation dose. Most importantly, it spreads out the breast tissue resulting in less superimposition and reduced likelihood of details being obscured by overlying tissue. By immobilising the breast, movement blur is minimised. Additionally, compression reduces overall breast thickness and makes tissue thickness more uniform across the breast. This results in lower radiation dose and less scattered radiation which would degrade the image contrast [51, 52]. Note that anti-scatter grids are also used to prevent scattered radiation from reaching the image but these increase the radiation dose required.

![Figure 2.4: Unfiltered molybdenum x-ray spectrum (dotted blue line) and effect of adding 0.03mm molybdenum filter (solid pink line). Generated using IPEM Report 78 Spectrum Processor [50]](image)

### 2.3 Anatomy of the breast

The internal anatomy of the breast is shown in Figure 2.5. It consists of fatty tissue, glandular components, the lymphatic network, blood vessels, connective tissue and supportive stroma (Cooper’s ligaments). The glandular tissue consists of 15 – 20 lobes that extend from the nipple in a radial pattern. Each lobe contains a tree-like pattern of ductal structures which are lined with two layers of epithelial cells. The most important part of the ductal structure is known as the terminal duct lobular unit (TDLU), which is responsible for milk production and hormonal and nutritional exchange. The number and size of the TDLUs will change with factors such as pregnancy, lactation and hormone fluctuation. Glandular tissue atrophies with age, starting in the medial posterior region and working its way to the nipple [51].
In order to fully visualise all tissue, the mammography examination comprises two complementary views of each breast. The cranio-caudal (CC) and medio-lateral oblique (MLO) projections are shown in Figure 2.6.

Breast anatomy is complex and it is extremely difficult to discern structures at an individual level on a mammogram, particularly the TDLUs. Instead, the radiological appearance of the structures within the breast is referred to as the parenchymal pattern, as shown in Figure 2.7. These are full-field digital mammography images taken on equipment from two different manufacturers.
Figure 2.7: Digital mammogram images with labelled components of the parenchyma

1 Thanks to Beverlee Macdonell-Scott (BreastScreen NSW) for assistance in selecting and annotating the images
In digital mammography, two images are available based on a single x-ray acquisition. The first is the raw (“for processing”) image; the second is the processed (“for presentation”) image which is used by the radiologists for diagnosis. The images in Figure 2.7 are therefore processed images. It is interesting to note the effect of image processing on the level of detail visible within the breast, as well as the different contrast. On occasion, it is possible to see the tubular appearance of individual blood vessels (Figure 2.7a) and the scalloped appearance of Cooper’s ligaments (Figure 2.7b).

The majority of studies examining the relationship between breast density and breast cancer risk assume that breast tissue is either adipose or glandular. The glandular tissue would therefore include the supportive and connective tissue (stroma) and the ductal structures, which contain the epithelium. However, some studies categorise breast density based on the complexities of the parenchymal pattern. The following sections describe classification schemes used to define mammographic density and how these have been employed in clinical studies. Some of the reasons why mammographic density is considered a risk factor are explained.

2.4 Breast density classification schemes

Classification schemes may be qualitative or quantitative. Studies using quantitative schemes, such as the Boyd Six Category Classification [14], are generally more successful in demonstrating a strong association between density and risk as these schemes are less subjective and therefore generate more reproducible results. However, qualitative schemes may be more effective in informing whether a particular woman with a high-risk parenchymal pattern would benefit from follow-up examinations such as ultrasound or tomosynthesis. There is currently interest in using digital image analysis to investigate the relationship between parenchymal pattern, as opposed to quantity of density, with risk. An EU grant has been approved and work will commence in 2013.

2.4.1 Wolfe Classification Scheme

Wolfe [12] described the breast parenchyma as having three components: fat, connective and epithelial tissues, and “prominent ducts”. The connective and epithelial tissues are grouped together under the term mammary dysplasia. The Wolfe Categories are as follows:

N1 (Lowest Risk): Primarily fat with small or no amounts of dysplasia.

P1 (Low Risk): Chiefly fat with prominent ducts in the anterior portion of the breast occupying no more than one quarter of the breast volume, or a band of ducts extending into a quadrant.

P2 (High Risk): Severe involvement with prominent duct occupying more than one quarter of the breast volume.

DY (Highest Risk): Severe involvement with dysplasia obscuring underlying prominent duct.
Two studies [12, 13] were carried out to evaluate the use of the Wolfe Classification Scheme. All subjects were aged over 30 years with an initial report negative for carcinoma and no history of breast cancer. A positive case was assigned if breast cancer was proven histologically at least 6 months after the radiographic examination. Although the studies differed slightly in their follow-up periods, the results were similar.

The carcinoma incidence in DY subjects compared to the incidence in N1 subjects was 37:1 for one study [12] and 22:1 for the other [13], suggesting that breast density is a particularly strong indicator for risk. Analysing the results of study [13] further, Wolfe found that the (N1 + P1) group comprised 67% of the population but contained only 22% of carcinomas and the (P2 + DY) group comprised 33% of the population but contained 78% of carcinomas. Splitting the study cohort into two groups rather than four, the relative risk was found to be less dramatic at 6:1 for carcinoma incidence in (P2 + DY) compared with (N1 + P1). Given that there were no established breast cancer screening programmes in 1976, Wolfe believed these groups could be used as the basis for routine screening and follow-up. Over 35 years later, this approach is being considered [11]!

Wolfe’s theory prompted several independent groups to conduct their own clinical studies in order to verify or reproduce his results; some studies were unable to do so. Boyd et al (1982) [16] believed these studies to be biased and undertook their own case-control study in an endeavour to validate Wolfe’s theory. Three groups of subjects were considered: a control group from a screening centre, a control group from a diagnostic referral centre (symptomatic controls) and a case group who had histologically verified breast cancer. Although the symptomatic controls had suspected abnormalities on physical examination, they were considered to be free of breast cancer after evaluation. There were 549 mammograms in total which were randomised and independently classified by three radiologists at different institutions with no knowledge of cases and controls. The mammogram of the non-cancerous breast was selected from the cases.

All radiologists identified the DY pattern more frequently in cases than ‘screened controls’. Odds ratios (OR) of 3.30, 1.88 and 3.67 were calculated for breast cancer risk associated with the DY category compared to N1, all of which were statistically significant. However, the number of women assigned to each category was very similar when comparing ‘symptomatic controls’ and there was no increase in relative risk between the DY category compared to N1. The reason for such a high prevalence of DY pattern amongst ‘symptomatic controls’ was the presence of a breast lump in 70% of the subjects, a symptom commonly associated with mammary dysplasia.

This study demonstrated that an association between the DY pattern and breast cancer is critically dependent on the selection of the controls. Using a control group where symptoms are common introduces referral bias and is likely to result in an underestimation of the risk associated with DY pattern. Such a control group was frequently selected in the studies that disproved Wolfe’s theory.
An additional review of 17 studies using Wolfe’s classification scheme was published by Boyd et al (1984) [53], with the aim of explaining contradictory results in the literature for breast cancer risk associated with the DY pattern. The authors defined a number of methodological standards expected of any epidemiological investigation of risk, such as population assembly (selection criteria, avoidance of referral bias), classification of mammographic pattern (observer variation, unbiased reading), analysis of breast cancer risk (effects of age and other risk factors accounted for) and follow-up (cohort studies only). It was generally found that those studies reporting an association between increased risk and DY pattern met more standards than those that found a negative link and therefore controversy in the literature can be attributed to methodological differences in the studies.

2.4.2 Boyd Classification Scheme

Although Boyd et al [16, 53] supported Wolfe’s theory of an association between DY pattern and breast cancer risk, they felt that there were limitations with such a subjective scheme and sought a more explicit categorisation of dysplasia than simply defining it as “severe”. The Boyd Six Category Classification (SCC) scheme was devised, based on the proportion of the breast occupied by mammographically dense tissue [14]. The six categories, shown in Figure 1.1 (Chapter 1), are 0%, < 10%, 10% to < 25%, 25% to < 50%, 50% to < 75% and ≥ 75%.

A case-control study [14] was carried out with assessment performed using both the Wolfe and the Boyd schemes. In this particular study, the 0% and < 10% categories were combined. Cases were subjects who had had a mammogram at or immediately before diagnosis and had uni-lateral disease. The mammogram of the non-cancerous breast was chosen to avoid reader bias. Controls were volunteers from a feasibility study for breast cancer screening and were all physically asymptomatic and confirmed to have no evidence of breast cancer after their mammogram. They had no personal history of breast cancer. A total of 183 case-control pairs were selected, matched to within 5 years of age, same side mammogram and as closely as possible to the year of mammogram. Mammograms were randomised and assessed independently by three radiologists at different institutions.

Using the Boyd scheme, it was found that many more cases than controls were assigned to the top two categories. A marked difference was observed in the highest category, which was occupied by 17% of cases and only 4% of controls. The OR calculated for the top category, ≥75%, with reference to the < 10% category was 5.99, 2.82 and 3.74 with statistical significance being achieved by all radiologists.

Statistical methods were applied to find out which classification scheme could better discriminate cases from controls. For two out of three radiologists it was found that extent of mammographic density alone (i.e. Boyd) was more effective; for the third radiologist both methods were found to
be equally effective. It is interesting to note that less observer variation was found when using the Wolfe scheme.

2.4.3 BI-RADS Lexicon

BI-RADS is the Breast Imaging Reporting and Data System established by the American College of Radiology (ACR) [28]. The system is a quality assurance tool designed to standardise mammographic reporting and facilitate outcome monitoring. Breast radiologists working in the USA must assign an assessment score to the examination, which dictates follow-up action. They also classify each mammogram into one of the following breast composition categories:

1. **The breast is almost entirely fat (<25% glandular)**
2. **There are scattered fibroglandular densities (approximately 25-50% glandular)**
3. **The breast tissue is heterogeneously dense, which could obscure detection of small masses (approximately 51-75% glandular)**
4. **The breast tissue is extremely dense. This may lower the sensitivity of mammography (>75% glandular).**

It is now a legal requirement in three US States (Connecticut, Texas and Virginia) that women are informed of their breast density when they receive their mammogram report. Even a federal bill is under consideration, which proposes that all women are informed of the breast cancer risk associated with their breast density and that women with high breast density are advised to discuss supplemental tests with their physician [54]. This appears to be the result of campaigns by the charity *Are You Dense, Inc* [55] who also established an educational website to promote awareness of mammographic density and the associated risks.

2.4.4 Tabár Patterns

The work by Tabár is reflective of that by Wolfe [12], since the mammogram is classified into patterns based on the parenchyma, rather than the proportion of mammographic density. The classification scheme is based on anatomic-mammographic correlations, following three-dimensional, thick-slice, histopathologic-mammographic comparisons [56]. The five patterns are:

1. **Mammogram composed of scalloped contours with some lucent areas of fatty replacement, and 1 mm evenly distributed nodular densities**
2. **Mammogram composed almost entirely of lucent areas of fatty replacement, and 1 mm evenly distributed nodular densities**
3. **Prominent ducts in the retroareolar area**
4. **Extensive nodular and linear densities, with nodular size larger than normal lobules**
5. **Homogeneous, ground glass-like appearance with no perceptible features**
Patterns IV and V are considered high-risk patterns and have been shown to exhibit strong correlations with other breast cancer risk factors, particularly reproductive factors [57, 58]. Pattern IV has also been shown to be associated with breast cancer risk in a case-control study, with an OR of 2.59 compared to the other patterns [59]. However, literature on the relationship between the patterns and breast cancer risk is sparse and it is thought that these patterns are better used as a teaching aid and to guide decisions on follow-up examinations.

2.5 Clinical studies of breast density

The following two case-control studies were carried out in parallel on the same study sample. Using 354 case-control pairs selected from a cohort of 45,000 from the Canadian National Breast Screening Study (NBSS), Boyd et al (1995) [21] applied the Boyd SCC scheme with density assessed visually and using an interactive thresholding technique [60]. Byng et al [22] applied techniques involving histogram skewness and fractal dimension analysis [61], believing this to be the first study to use a fully automated technique to equate density with risk. These quantitative techniques are described in Chapter 3.

Cases were women in whom histologically verified breast cancer developed at least 12 months after entry to the trial. This period was chosen to remove any masking bias. Controls were matched to each case by year of entry into the NBSS, age at entry (± 1 year), and the centre at which the case was diagnosed. The length of time the control subject was enrolled in the study must have exceeded the period of time the case was in the NBSS before cancer was diagnosed. The mammogram taken at entry to the NBSS was used for classification; the CC view of the cancer-free breast was selected for the case and the corresponding view for the controls. One view was considered adequate as the authors found a high degree of symmetry between left/right and CC/MLO breasts.

For the group as a whole, Boyd et al [21] observed increases of 43% and 32% in the relative risk of breast cancer for each density category, based on radiologist visual assessment and interactive thresholding respectively. Subjects in the category of most extensive density (≥ 75%) had a relative risk of 6.05 compared to subjects in the category with zero density when using radiologist classification, and a relative risk of 4.04 when using interactive thresholding. Risk estimates for younger women (40 – 49) were higher that those for older women (50 – 59) using both techniques.

Byng et al [22] quoted a maximum relative risk of 3.65 using the histogram skewness measure and 2.86 for the fractal dimension analysis, both somewhat lower than the results from the more subjective SCC classification, but still demonstrating a strong association between mammographic density and risk.
Boyd et al (2007) [18] published the results of three nested case-control studies with 1,111 matched case-control pairs. Women were recruited from the NBSS, the Screening Mammography Programme of British Columbia and the Ontario Breast Screening Programme. Inclusion and matching criteria were identical to those described above. Again, the CC view from the cancer-free breast of the baseline mammogram was used for the cases and the corresponding view for the controls. Density was estimated using the interactive thresholding technique and assessed by two radiologists.

In order to examine the effect of masking bias, the study included both screen-detected and interval cancers, although subjects who had a diagnosis of breast cancer less than 12 months after their first screening mammogram were excluded.

Subjects with $\geq 75\%$ density were compared to those with $<10\%$ density. Considering all cancers detected by any means, the Odds Ratio (OR) was 4.7 (95% confidence interval (CI) 3.0 to 7.4); for cancers detected by screening only, the OR was 3.5 (95% CI 2.0 to 6.2); for cancers detected by physical means less than 12 months after the last screening exam, the OR was 17.8 (95% CI 4.8 to 65.9) and for those detected more than 12 months after the last screening exam, the OR was 5.7 (95% CI 2.1 to 15.5). Furthermore, the elevated risk was found to persist for at least 8 years after study entry. 16% of all cancers, 12% of screen-detected cancers and 40% of cancers detected less than 12 months after the last negative screening mammogram were found in breasts with $\geq 50\%$ mammographic density. This percentage was even higher in women below the median age of 56, as was the increased risk.

Despite the differences in the estimated value of relative risk, these studies [18, 21, 22] have shown that regardless of the method of assessment, mammographic density is a strong risk factor, and remains so even after accounting for other confounding risk factors. The method of cancer detection in cases has a significant impact due to the masking effect and it is therefore thought that the best estimate of overall risk is calculated by combining cancers detected by screening with those detected up to 12 months after the last screening examination.

2.6 Possible mechanisms of breast cancer risk associated with breast density

A ‘breast density gene’ has not yet been identified and the current hypothesis for the link between density and risk is due to the combined effects of cell proliferation (mitogenesis) and damage to the DNA of dividing cells (mutagenesis, by products of lipid perioxidation) [62, 63]. These are not independent processes but have an additive effect on each other, with cell proliferation promoting lipid perioxidation, and vice-versa. Urinary levels of mutagenic products of lipid perioxidation are linked to both the risk of breast cancer and mammographic density [62, 63].
Increased cell proliferation is observed in the epithelium and stroma in response to growth factors and hormone exposure [63]. The blood concentrations of Insulin Growth Factor IGF-I in pre-menopausal women and prolactin in post-menopausal women are associated with increased mammographic density and an elevated risk of breast cancer [64]. Additionally, the tissue surrounding breast lesions in women with extremely dense breasts has been found to show increased IGF-I, greater nuclear area and more total collagen [65]. Cumulative lifetime exposure to oestrogen is also a contributing factor, with use of hormone therapy, parity and age at menopause all showing an association with both mammographic density and breast cancer risk. These topics are discussed further in the next section.

2.7 Breast cancer risk factors

Breast cancer risk factors are summarised in Table 2.2 [66, 67] and discussed in the following paragraphs. Their relationship with mammographic density is described, where appropriate.

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Relative risk</th>
<th>High-risk group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>&gt; 10</td>
<td>Elderly individuals</td>
</tr>
<tr>
<td>Geographical location</td>
<td>5</td>
<td>Developed countries</td>
</tr>
<tr>
<td>Breast density</td>
<td>&gt; 5</td>
<td>Extensive mammographic density</td>
</tr>
<tr>
<td>Previous benign breast disease</td>
<td>4 – 5</td>
<td>Atypical hyperplasia</td>
</tr>
<tr>
<td>Cancer in other breast</td>
<td>&gt; 4</td>
<td>Previous breast cancer</td>
</tr>
<tr>
<td>Exposure to ionising radiation</td>
<td>3</td>
<td>Abnormal exposure to young girls after age 10 years</td>
</tr>
<tr>
<td>Age at menarche</td>
<td>3</td>
<td>Before age 11 years</td>
</tr>
<tr>
<td>Age at first full pregnancy</td>
<td>3</td>
<td>First child after age 40 years</td>
</tr>
<tr>
<td>Age at menopause</td>
<td>2</td>
<td>After age 54 years</td>
</tr>
<tr>
<td>Family history</td>
<td>≥ 2</td>
<td>Breast cancer in first-degree relative</td>
</tr>
<tr>
<td>Socioeconomic group</td>
<td>2</td>
<td>Groups I (high status) and II (low status)</td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.7</td>
<td>High body mass index</td>
</tr>
<tr>
<td>Pre-menopause</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Post-menopause</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hormone Replacement Therapy</td>
<td>1.66</td>
<td>Current users</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>1.07</td>
<td>7% increase with every daily drink</td>
</tr>
</tbody>
</table>

Breastfeeding and parity are protective factors with relative risk falling by 4.3% for every 12 months of breastfeeding, in addition to a 7% reduction for every birth [66].

Table 2.2: Established risk factors for breast cancer [66, 67]. Hormonal risk factors are highlighted
For women, the UK lifetime risk of developing breast cancer is currently 1 in 8 [10], although an individual’s risk is highly dependent on a number of factors. The increasing risk, and consequently incidence, over time is inevitably related to an ageing population. There is little doubt that environmental and lifestyle factors have also been hugely influential. This is reflected in the elevated risk in countries with traditionally very low incidence, such as Japan and the developing world, attributed to their adoption of a “Westernised” lifestyle, particularly in terms of diet and reproductive behaviour [68, 69]. In the UK, the population attributable fraction (PAF) for breast cancer is estimated as 26.8%; this represents the proportion of cancers linked to sub-optimal levels of major environmental and lifestyle factors, which could potentially be prevented [70].

2.7.1 Age

Breast cancer risk is strongly related to age, as shown in Table 2.3, with 81% of UK cases occurring in women aged 50 years and over [10]. Breast cancer incidence increases with age, doubling approximately every 10 years until the menopause, when the rate of increase slows dramatically [67]. This is contrary to most cancers, which exhibit a linear relationship between incidence and age when both variables are plotted on a logarithmic scale [71].

<table>
<thead>
<tr>
<th>Estimated risk at birth up to and including:</th>
<th>UK (2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>age 29</td>
<td>1 in 2,000</td>
</tr>
<tr>
<td>age 39</td>
<td>1 in 215</td>
</tr>
<tr>
<td>age 49</td>
<td>1 in 50</td>
</tr>
<tr>
<td>age 59</td>
<td>1 in 22</td>
</tr>
<tr>
<td>age 69</td>
<td>1 in 13</td>
</tr>
<tr>
<td>Lifetime risk</td>
<td>1 in 8</td>
</tr>
</tbody>
</table>

*Table 2.3: Age-specific risk of breast cancer [10]*

Breast cancer risk is therefore strongly dependent on hormonal risk factors. These are accounted for by the Pike model [72] which proposes that breast tissue age, rather than chronological age, is the relevant measure for describing the age-specific incidence of breast cancer [64, 72].

The Pike model is shown in Figure 2.8 (A). The points on graph (B) are the age-specific breast cancer incidence rates in US white women taken from the Third National Cancer Survey, 1969 – 71 [73] and the solid line is the fitted curve from the Pike model, which shows excellent agreement.
Figure 2.8: Pike model of breast tissue ageing (A) and log-log plot of age specific incidence of breast cancer in the USA, 1969 – 71, with fitted curve from Pike model (B)
Reproduced with permission from [64], RightsLink Copyright Clearance Center

The model proposes that breast-tissue ageing starts at menarche and continues at a rate $f_0$ until first full term pregnancy (FFTP). It then slows to rate $f_1$, decreases further in the peri-menopausal period and reaches a minimum rate, $f_2$, at menopause. A constant, $b$, is incorporated for women with late FFTP, as they are considered to be at higher risk of developing breast cancer than nulliparous women (see Section 2.7.3).

Mammographic density diminishes with increasing age, with a noticeable decline observed after the menopause [62]. It therefore appears paradoxical that breast cancer incidence increases with age and is more common in post-menopausal women. However, mammographic density shares many features with breast-tissue age suggesting that cumulative exposure to mammographic density could therefore reflect cumulative exposure to hormones and growth factors that stimulate cell division in the breast stroma and epithelium. This exposure could be an important factor underlying the age-specific incidence of breast cancer [64].

2.7.2 Menarche and menopause

Menarche is defined as the start of menstrual periods and menopause is the complete cessation of menstrual periods. Menopause may be natural, whereby ovaries naturally decrease their production of oestrogen and progesterone and there are no menstrual periods for 12 consecutive months, or induced, which indicates surgical removal of the ovaries (bilateral oophorectomy) [74]. Women with an early age of menarche (< 11 years) or late natural menopause (> 55 years) have an increased risk of developing breast cancer, as both result in a longer exposure to oestrogen.
and progesterone. Furthermore, women who have an induced menopause before the age of 35 have only 40% of the risk of breast cancer of women who have a natural menopause [67].

The menopause is associated with a decrease in mammographic density; studies have shown both a reduction in the area of dense tissue on the mammogram and an increase in the area of non-dense tissue [62, 64]. There is little in the literature regarding the correlation between age at menarche and breast density, but it is thought that mammographic density may be a marker of breast-tissue ageing. Based on the Pike model, it would therefore be expected that early menarche would be associated with an increase in mammographic density.

This hypothesis was examined in a recent UK longitudinal study which found that some characteristics of mammographic density make it a suitable marker of breast-tissue ageing in relation to the Pike model [75]. For example, it was found that percentage mammographic density declined over the age range of 50 – 64 and dropped by 2.4% at menopause transition. Effects of parity and menopause were also consistent with the predicted determinants in the Pike model. However, whereas the model assumes a constant rate of tissue ageing after menopause, the study found that annual rates of decline in breast density slowed with age (2.4% at age 50, 0.7% at 57 and 0.1% at 64).

A possible confounding factor in studies investigating change in mammographic density over time is the phase of the menstrual cycle at the time of mammogram. The mitotic rate of breast cells is 2 – 3 times higher during the luteal phase of the menstrual cycle than during the follicular phase [76] and this is manifested in mammographic density. Studies have shown that PMD during the luteal phase is up to 7% higher than that in the follicular phase [77, 78] although this was not statistically significant. However, these studies had very small populations of 11 women [77] and 36 women [78]. Results also showed that in some women, there was no change, or even a decrease in PMD between the luteal and follicular phases [77] and that the PMD on two mammograms both taken in the same phase, 12 months apart, could increase by up to 3% [78].

Without further evidence to the contrary, it does not seem necessary to account for cyclical variations in studies on mammographic density. Furthermore, the vast majority of women attending for breast screening will be post-menopausal.

2.7.3 Parity and age at first birth

Early age at first term birth and increased parity are associated with a lifetime reduction in breast cancer risk. A women with no children (nulliparous) has roughly the same risk as a women with a first term birth aged about 30 years and the highest risk group is those who have their first child after the age of 35 [66, 67, 79]. Risk is additionally reduced by breastfeeding for a period of 12 months or more [66].
Woolcott et al [80] analysed data from four case-control studies carried out in the USA and Japan to examine the effects of parity and age at first birth on mammographic density and additionally, to determine if the association between mammographic density and risk was modified by reproductive factors. Mammographic density (both PMD and dense area) was lower in parous than nulliparous women and decreased with greater parity; it was lowest in women with 3 or more children. However, within the parous group, it did not vary by age at first birth. When considering the populations from all four studies as a single group, the association between density and risk appeared to be stronger among nulliparous than parous women. However, further analyses showed that this was only true for PMD, not dense area. There were no significant differences in the association between density and risk with age at first birth and increased parity. The authors concluded that mammographic density explains only a very small proportion of the reduction in breast cancer risk associated with parity, and that the two can be considered as independent risk factors.

2.7.4 Weight / Body Mass Index (BMI)

Weight and BMI show a strong positive association with breast cancer in post-menopausal women, but an inverse relationship is actually observed in pre-menopausal women [66, 67]. The association is likely to be hormone-related as adipose tissue influences exposure to oestrogen, especially in post-menopausal women [81].

The correlation with breast cancer risk may not be as straightforward as weight at a single point in time, but related to adult weight gain over specific lifetime periods. Harvie et al [82] examined weight change over the periods 18 – 30 years, 30 years to menopause and post-menopause, in a population of 33,660 US women. The most common trend was that women gained weight consistently throughout their adult life; these women were classed as the reference group and had the highest rates of post-menopausal breast cancer. Pre-menopausal loss of weight reduced the breast cancer risk relative to the reference group and also relative to those who maintained their weight. Weight gain after age 30 was more detrimental than weight gain from 18 – 30 years, but results were not statistically significant. Weight gained during this period is more likely to manifest as abdominal fat but studies have not shown an association between waist-hip ratio, or waist circumference, and breast cancer risk, after adjustment for BMI. Additionally, results on the relationship between dietary fat intake and risk are inconsistent [66, 67].

Numerous studies have shown that weight and BMI are inversely associated with PMD [62]. In a study involving 1,114 matched case-control pairs, Boyd et al [81] found significant correlations between BMI with PMD ($r = -0.41$, $p < 0.0001$), dense area ($r = -0.10$, $p < 0.01$), non-dense area ($r = 0.59$, $p < 0.0001$) and total breast area ($r = 0.56$, $p < 0.0001$). Similar results were observed for weight but the strength of correlation was slightly lower. Originally, BMI showed no statistically significant associations with breast cancer risk but, after adjustment for PMD, the association was
positive and significant for all subjects, including pre-menopausal women. Additionally, the risk associated with PMD was found to increase after adjustment for BMI.

BMI and PMD must therefore be treated as independent risk factors, operating through different pathways. Since each acts as a confounder for the other, it is essential that this is controlled for, otherwise the effects of each variable on the risk of breast cancer will be underestimated.

2.7.5 Hormone therapy

Errors can occur during the DNA replication phase of cell division. Genetic mutations, and potentially cancer induction, are the consequences of those errors which are not repaired. Oestrogens are known to increase the rate of cell proliferation in glandular tissue and therefore amplify the risk of initiation and promotion of breast cancer. Progestins are also thought to enhance the rate of breast cell proliferation [83].

In post-menopausal women, a number of studies on the effects of hormone replacement therapy (HRT) have shown that the increase in mammographic density is greatest in users of oestrogen plus continuous progesterone (EPT), followed by oestrogen plus cyclic progesterone, then oestrogen-alone (ET); the relative risk of breast cancer also follows this trend [83, 84]. The **Women’s Health Initiative** study randomly allocated 16,000 post-menopausal women to receive either EPT or placebo. After 5 years of follow-up, the investigators reported a 26% increase in the relative risk of breast cancer with EPT compared to placebo and the study was terminated early for ethical reasons [85]. Breast proliferation with EPT is localised to the TDLUs, which are the site of development of most breast cancers.

Although an increase in mammographic density is observed during use of HRT, this is not sustained following discontinuation. Rutter et al [86] compared the BI-RADS density [28] on mammograms of 5,212 post-menopausal women taken at two screening appointments 11 – 25 months apart. Women were placed into 4 groups depending on their HRT use at the time of both exams: non-users, discontinuers, initiators and continuing users. Compared to the non-user group, the initiators showed an increase in density, the discontinuers showed a reduction in density and the continuing users exhibited higher density at both exams. A separate study also found that cessation of HRT is followed by a rapid decrease in density, observable within just 2 weeks [65].

Given the relationships discussed above, mammographic density could be considered a surrogate marker for the effects of hormone therapy on breast cancer risk. Boyd et al [87] rejected this hypothesis after examining PMD and HRT use within 724 matched case-control pairs. As expected, density was greater in cases than controls and greater in current users of HRT compared to past or never-users. The relative risk of breast cancer was found to increase with
HRT use, when considered both with and without adjustment for PMD. The authors therefore concluded that the effects of HRT use on density and risk are separate and not causally related, although they acknowledge some limitations with the study, one of which being that no attempt was made to distinguish whether ET or EPT was being used. A more suitable marker may in fact be change in mammographic density, as this is related to breast cell mitotic activity [88, 89].

Selective oestrogen receptor modulators, such as Tamoxifen and Raloxifene demonstrate the opposite effects of HRT; that is, they decrease mammographic density and are associated with decreased breast cancer risk [65]. Cuzick et al [90] have recently shown that the extent of reduction in mammographic density over a 12 – 18 month period of taking tamoxifen is an excellent predictor of how successful the treatment will be. In a randomised prevention trial of tamoxifen versus placebo, it was found that women who experienced a reduction in mammographic density of at least 10% (based on visual assessment) had a 63% reduction in breast cancer risk compared to the placebo group. Women with less than 10% reduction in density had no risk reduction, suggesting that these women would not benefit from tamoxifen and should seek alternative risk-reduction strategies.

2.7.6 Family history and genetics

Genetic predisposition is thought to account for up to 10% of breast cancer in Western countries [67], with the two highest-risk genes, BRCA1 and BRCA2 accounting for 4 – 5% of cases overall [91] and a substantially higher percentage (at least 20%) in women below the age of 30 [92]. Other identified genes include p53 and PTEN, but the total number remains to be determined.

Family history is also a risk factor for breast cancer, particularly for those women with at least one first degree relative (mother, sister or daughter) who developed the disease. The risk increases with the number of affected relatives and is greater if they developed the disease below the age of 50 [67]. Family history of the disease is not necessarily affiliated with a BRCA1 or BRCA2 mutation. Metcalfe et al [93] showed that women with a significant family history of breast cancer (two or more breast cancers under the age of 50 years, or three breast cancers at any age) who tested negative for BRCA mutations, still had an approximately four-fold risk of breast cancer.

There remains a lack of evidence regarding whether a family history of breast cancer affects mammographic density [62]. However, mammographic density is strongly associated with breast cancer risk among women with a family history of the disease and has been shown to have an additive effect. In a case-control study, Boyd et al [94] found that the relative risk of breast cancer for women with ≥ 75% PMD, compared to those with < 10% PMD was 11.14 amongst subjects with at least one first degree relative.
Mammographic density is likely to be inherited although the genes responsible, and the role they may have in causing breast cancer, are currently unknown. A recent study on 571 pairs of monozygotic twins (genetic copies of each other) and 380 pairs of dizygotic twins (sharing about half their genes) in Australia and North America concluded that heritability accounted for 63% of the variation in density. PMD was measured on one cranio-caudal view for each woman and the study participants provided consent to release their most recent mammogram for the purpose of density assessment. If the mammogram was more than 2 years old, they were asked to make an appointment with their local screening programme. The study required that mammograms in each pair of twins were obtained within 36 months of each other. After adjustment for age, reproductive factors and BMI, the correlation of PMD between twins was 0.27 for dizygotic pairs and 0.63 for monozygotic pairs, which is consistent with an additive genetic cause. Results were very similar between US and Australian populations [15].

2.7.7 Benign breast disease

Examples of benign breast disease include palpable cysts, fibroadenomas and duct papillomas; these are all associated with a modest increase in the risk of developing breast cancer. However, women with severe atypical hyperplasia are at approximately four times higher risk than those with no proliferative changes in their breast [67]. Literature on the relationship between benign breast disease and mammographic density is currently scarce and contradictory [65].

2.7.8 Ethnicity, geographical variation and socioeconomic status

Age-standardised incidence and mortality rates for breast cancer throughout the world are shown in Figure 2.9 [2]. The data are taken from the GLOBOCAN project, undertaken by the World Health Organisation (WHO) and reflect the female population, not just those women who have attended for breast screening. Cancer incidence data have been derived from population-based cancer registries, which may also produce survival statistics. However, the mortality statistics presented in Figure 2.9 were collected and disseminated by the WHO. The confidence in the data is high for European and American countries but lower for African and Asian countries, which do not have comprehensive death registration systems, Age-standardisation is necessary when comparing several populations that differ with respect to age and life-expectancy. The age-standardised rate is a weighted mean of the age-specific rates. In this case, the weights are taken from the population distribution of the GLOBOCAN world standard population [2].

Incidence rates are typically much higher for Western countries, although the difference between Eastern and Western populations has decreased in recent years [68, 69]. Furthermore, the breast cancer risk in Asians who migrated to the USA has been found to approach the risk of US Caucasians after two or three generations [67]. This suggests that lifestyle and environmental factors may be a stronger determinant of breast cancer risk than ethnicity.
However, a variation in mammographic density with ethnicity persists after accounting for covariates, in particular age and BMI [95, 96]. Maskarinec et al [95] examined the relationship between mammographic density and breast cancer incidence in 1,327 women from four ethnic backgrounds (Japanese, Hawaiian, Caucasian and Latina) living in four locations (Japan, Hawaii, Arizona and Norway), including both the indigenous population and migrants. Age-adjusted dense area had the strongest association with breast cancer incidence ($r = 0.93; p = 0.03$), closely followed by dense area adjusted for age, BMI and parity ($r = 0.86, p = 0.01$). Percent mammographic density (PMD) was poorly correlated with incidence when adjusted for age only, but the correlation improved when PMD was further adjusted for BMI and parity. These results support the hypothesis that density is related to genetic factors.

McCormack et al [96] compared the mammographic density of first generation South-Asian and Afro-Caribbean women to Caucasian women in the UK. These ethnic minority groups have a 20 – 30% lower incidence of breast cancer. Age-adjusted PMD was found to be 5.6% and 5.9% lower in Afro-Caribbean and South-Asian women respectively, reducing to 1.3% and 3.8% lower after adjustment for all confounding factors as these groups typically exhibited higher BMI, a more protective reproductive history and lower use of hormones (HRT and oral contraceptive). The differences in absolute dense area were not significant between Afro-Caribbeans and Caucasians.
but followed similar trends to PMD for the South-Asian cohort. The mean age of migrants was less than 30; no correlation was observed between density and number of years spent in the UK.

An additional finding in both developed and developing countries is that higher socioeconomic status (SES) is associated with increased breast cancer incidence although whether this relationship is independent of known breast cancer risk factors remains controversial because SES variables are likely to be markers for lifestyle and anthropometric variables as well as uptake of breast screening. SES is typically assessed by one or more of the following variables: income, occupation, level of education and area of residence.

Aitken et al [97] examined the variation in mammographic density by education, deprivation score and urban/rural residence in a population of 487 pre-menopausal women. The study population was skewed towards more affluent areas and higher levels of education. Education was assigned to one of 3 groups: up to GCSE, up to A-level and university-level. PMD was found to be 6.3% higher in the most educated group compared to the least educated but the difference reduced to 4.0% after adjusting for BMI. Other risk factors were not confounders. Deprivation score ranged from 1 (“deprived”) to 5 (“affluent”) based on standardised scores of variables such as unemployment and home-ownership, taken from Census data. Unadjusted PMD was 6.6% higher in the most affluent area compared to the least affluent area but this was reduced to -0.6% when adjusted for BMI. There was no difference in PMD between urban (>10,000 people) and rural (<10,000 people) populations. Absolute dense area showed no correlation with SES for any of the groups. The results indicate that mammographic density does not fully explain the SES risk gradients and that BMI is a strong confounding factor.

2.7.9 Use of underarm products (deodorants and antiperspirants)

The relationship between deodorant / antiperspirant use and breast cancer risk remains controversial but the hypothesis stems from the fact that a high proportion of breast carcinomas arise in the upper outer quadrant (UOQ). This is attributed to a greater amount of breast (epithelial) tissue in this region, a finding supported by biopsy studies [98]. However, the recorded quadrant incidence of breast cancers suggests that there has been a disproportionate increase in the UOQ over time. Early US studies showed the proportion to be approximately 31% (in 1926) but by 1994, the proportion had increased to 61% [99]. Results from the UK are less dramatic but show an increase in the proportion of UOQ cancers from 47.9% in 1979 to 53.3% in 2000 [100].

The proposed mechanism is attributed to aluminium and parabens. Aluminium salts are the active ingredient in antiperspirants as they block the sweat glands. It is thought that these can cause direct alterations to DNA. Aluminium is also considered a metalloestrogen which means it may interfere with oestrogen in human breast cancer cells [99, 101]. Parabens (esters of p-hydroxybenzoic acid) are anti-microbial preservatives present in both antiperspirants and deodorants. As a group they have been shown to exhibit low toxicity but they may have
oestrogenic and endocrine disrupting effects [102]. Animal studies lasting 3-4 days showed that parabens are weak oestrogens and several hundred mg per kg body weight are required to produce observable effects. However, there is a lack of evidence on their long term effects in humans.

Underarm hair removal (shaving in particular) prior to the application of antiperspirants or deodorants may raise the risk as small nicks or abrasions could increase the likelihood of substances being absorbed. Mirick et al [103] carried out a retrospective analysis of 1,600 women (800 cases, 800 controls matched only by age) using a questionnaire which asked whether they used antiperspirant or deodorant, questions pertaining to hair removal (which method and how frequently) and whether antiperspirant or deodorant was frequently applied within one hour of shaving. They reported no increase in risk with any of these activities. Antiperspirant / deodorant use is not believed to impact upon mammographic density. A recent study looking at the spatial distribution of mammographically dense tissue within the breast found that it was typically clustered in the central region, not in the UOQ [104].

2.7.10 Alcohol and smoking

Some studies have shown a weak positive association between alcohol intake and breast cancer incidence, but other dietary factors may be covariates [66, 67]. Smoking is unconfirmed as a risk factor for breast cancer. A plausible positive association is indicated by a higher prevalence of p53 gene mutations (which are found in 15 – 30% of breast cancers) present in the breast tissue of smokers. However, smoking has demonstrated an anti-oestrogenic effect, suggesting a negative association. The fact that smoking exhibits opposing carcinogenic and anti-oestrogenic properties may explain the inconsistency in the literature. Terry and Rohan [105] reviewed 228 articles and concluded that recent epidemiological studies do not support a decrease in breast cancer risk with smoking but there may be an increase in risk with long duration and smoking before first full-term pregnancy.

Mammographic density appears to be positively correlated with alcohol consumption, although this may be dependent on the beverage [65]. Cabanes et al [106] found a significant positive association between current alcohol intake and PMD among post-menopausal non-smokers. Current smokers were found to have lower PMD than non-smokers, with the trend reflected in the number of cigarettes smoked daily. Two further studies [107, 108] also confirmed lower PMD (and dense area [107]) in peri- and post-menopausal current smokers and observed a significant inverse dose-response relationship between the number of cigarettes smoked and PMD. All three studies concluded that the negative association between PMD and smoking supports the anti-oestrogenic effect of smoking on breast tissue [106 - 108].
2.8 Breast cancer risk prediction models

Established mathematical models exist for the estimation of individual risk, including the Gail [23], Claus [24] and Ford [25, 26] models. However, they are limited in that they do not integrate information on family history, hormones and benign breast disease in a comprehensive fashion [109]. Tyrer et al [110] sought to overcome these limitations by developing a model which incorporates risk factors related to hormonal and reproductive exposures, family history and personal and medical history. Particular advances on previous work were the inclusion of a hypothetical gene that acts as a surrogate for the effects of all unknown genes, since mutations of the BRCA1 and BRCA2 genes alone cannot account for the increased risk associated with family history. It is thought that these genes only account for a maximum of 5% of all cases of breast cancer [64, 91, 110]. Risk factors such as age at menarche, parity, age at first childbirth, age at menopause, atypical hyperplasia, lobular carcinoma in-situ, height and Body Mass Index (BMI) were also added to the model, but not breast density.

Recently, Barlow et al [27] developed a model including mammographic density and the use of hormone therapy as additional risk factors. Density was classified by use of the BI-RADS categories [28]. They found that mammographic density was a statistically significant risk factor for breast cancer diagnosis in pre- and post-menopausal women. Hormone therapy was only statistically significant in post-menopausal women. The authors believe that the addition of density as a risk factor may identify high-risk women better than the Gail model. Chen et al [111] also found that the addition of breast density to the Gail model predicted higher risks for women with a high percentage of dense area.

The models discussed above have shown great accuracy in predicting the number of cancers within a population, but their ability to provide accurate individual estimates of absolute risk remains limited [17]. This issue is currently being addressed in a Manchester-based study entitled Breast cancer risk assessment and validation in the National Breast Screening Programme, referred to as PROCAS (Predicting Risk Of Cancer At Screening) [112, 113].

2.9 Discussion

In this chapter, mammographic density has been shown to be an important risk factor for breast cancer. It is strongly correlated with a number of other risk factors but after accounting for these confounding factors in clinical studies, association of mammographic density with breast cancer risk remains strong, making it an independent risk factor for the disease. It has also been shown to improve the predictive power of risk models, and may be used as a factor in determining tailored screening intervals based on individual risk. The next chapter describes methods developed to measure breast density by area and volume and examines how these will be affected by the transition from film-screen to digital mammography.
3. Quantitative Methods for the Determination of Breast Density

In the previous chapter, classification schemes for the categorisation of breast density were presented. This chapter examines quantitative methods for measuring breast density as a continuous variable. Both area-based and volumetric techniques are presented and the advantages and disadvantages of each method are discussed. Sections 3.1 and 3.2 explain how these methods are applied to digitised film-screen mammograms and Section 3.3 considers how they translate to digital mammography, which is rapidly replacing film-screen mammography in the NHSBSP.

3.1 Area-based methods

These include thresholding and segmentation techniques as well as texture-based methods, all of which require digitised images. Simpler techniques that can be used directly on film include planimetry and the visual analogue scale.

3.1.1 Planimetry

A transparent plastic sheet or acetate is placed over the mammogram and the breast area and regions of dense tissue are traced with an instrument known as a planimeter which integrates the enclosed area [65]. It was used in early studies by Wolfe [114] who found that intra-observer consistency was reasonable (±6% on 40 cases re-measured) and results for percentage mammographic density correlated well with the Wolfe categories. The most obvious drawback with planimetry is that it is labour-intensive, particularly if there are small isolated areas of dense tissue. Consequently, its use is not widely reported.

3.1.2 Interactive thresholding

In 1994, Byng et al [60] developed an interactive thresholding technique to quantify the percentage projected area of dense tissue on a mammographic image. Following the digitisation of a mammogram, the user identifies two threshold pixel intensities. The first of these, $i_{\text{edge}}$, separates the image of the breast from the background. Summing all pixels over intensity range $i_{\text{edge}}$ to $i_{\text{max}}$ (where $i_{\text{max}}$ corresponds to a white pixel) provides a measure of breast area. The second threshold, $i_{\text{DY}}$, identifies regions of dense tissue. Summing all pixels over the range $i_{\text{DY}}$ to $i_{\text{max}}$ gives the area of dense tissue. The percentage mammographic density (PMD) is then calculated as the ratio of the two. This process is shown in Figure 3.1.
The user interface of the current software version (known as Cumulus v4) is shown in Figure 3.2 [115]. Whilst viewing the mammogram, the user uses the sliding scale (top right) to adjust the threshold pixel intensity and a colour graphics overlay informs their decision as to when the optimal values have been set.

Cumulus is currently considered to be the gold standard [116 – 119]. In large-scale case-control studies [18, 21], it has consistently shown a strong correlation with breast cancer risk, with studies finding a similar magnitude of relative risk between the highest ($\geq$75%) and lowest (<10%) density.
categories. Using trained operators, the intra-observer agreement is extremely high, with reproducibility values of 0.94 [18] and 0.90 [21] reported within and between sets. Furthermore, it is not necessary for the operator to be a radiologist. Byng et al [60] used two radiologists, a clinical scientist and a research assistant to evaluate 70 images and found that the best intra-observer agreement \( r = 0.99 \) was demonstrated by the research assistant. The intra-class correlation coefficient for pairs of observers ranged from 0.84 – 0.94 indicating excellent inter-observer agreement.

The main limitation with Cumulus is that it requires “a major commitment of human resources” [116], despite evaluation times of “typically less than one minute per image” originally quoted by Byng et al [60]. Inevitably, there is a degree of subjectivity associated with interactive techniques and although this is low, a fully automated technique would be desirable.

3.1.3 Visual Analogue Scale

The reader is presented with a line (usually of length 10cm) for each mammographic view and simply places a mark on each line corresponding to their estimation of density in that image. The distance is measured and converted to the percentage mammographic density, a process which can be manual or automated. An example is shown in Figure 3.3.

```
RCC |---------------| |---------------|
RMLO| | | | |
| | | | |
LCC |---------------| |---------------|
LMLO| | | | |
```

*Figure 3.3: Schematic example of visual analogue scale

\[ RCC = 40\%; LCC = 42\%; RMLO = 43\%; LMLO = 45\% \]*

Duffy et al [60] have shown that this method of density assessment is strongly correlated with breast cancer risk when the average density from the MLO and CC view of a single breast is considered and categorised according to the Boyd Six Category Classification (SCC) scheme [14]. Mammographic density data was available for 10,048 women, including 311 cancers. The study population was taken from two screening centres and four radiologists at each centre shared the density reading. When only the MLO view was available, the odds ratio, OR, for the \( \geq 75\% \) category compared to the combined categories with <25\% density was only 1.51. However, when the average MLO and CC density was considered, the corresponding OR was 6.77. This is comparable to results found from studies using Cumulus [18, 21] although, interestingly, the Cumulus studies only used the CC view for density assessment. Unfortunately the intra- or inter-observer agreement is not reported in the visual assessment study but is assumed to be high to justify sharing the density reading.
Studies treating visually assessed density as a continuous variable are rare and it seems more common to assign breast density into one of the SCC [14] or BI-RADS categories [28]. Using BI-RADS, Ooms et al [120] found an intraclass correlation coefficient (ICC) of 0.77 (95% CI: 0.69 – 0.85) for four experienced radiologists. Ciatto et al [121] found that average interobserver agreement amongst 12 dedicated breast radiologists was moderate ($\kappa = 0.54$, range 0.02 – 0.77) although extremely poor agreement was found between two individual radiologists and the majority result ($\kappa = 0.02$, $\kappa = 0.03$).

Subjectivity is clearly a limitation of visual assessment. However, it has the major advantage of being extremely quick, adding only a few extra seconds to the reading time for a mammogram. Additionally, it does not treat the breast as a two-component model where “dense” tissue includes fibrous, glandular and other tissues. Radiologists may take the different structures into account when assigning visual density, which could explain the strong relationship with risk.

3.1.4 Automated techniques

A number of techniques of this nature exist but few have progressed beyond the research environment and therefore only a very brief summary is presented.

Zhou et al [122] carried out automatic segmentation using a gradient-based breast boundary tracking method to separate breast from background and using the characteristic shape of the grey-level histogram to segment the dense tissue from the non-dense, allowing images to be grouped into four classes, roughly corresponding to the BI-RADS categories [28].

Caldwell et al [123] developed an algorithm based on fractal dimension analysis [124] to characterise the parenchymal patterns of mammograms using the Wolfe scheme [12, 13] and compared the results to radiologist classification. Agreement between radiologist and fractal dimension analysis was 84% in calling the Wolfe grades, compared to radiologist inter-observer agreement of 85%.

Byng et al [61] combined elements of their thresholding technique [60] with the fractal concepts developed by Caldwell et al [123] and histogram analysis to distinguish density using both texture and brightness. A low density breast would have a coarse texture and a high fractal dimension due to good contrast between the predominantly fatty tissue (dark appearance) and the denser connective tissue (bright). Conversely, a highly dense breast would exhibit a low fractal dimension due to the smoother texture caused by low contrast between regions of similarly dense tissue. This was combined with analysis of the pixel intensity histogram where the shape of the histogram, in particular the asymmetry (skewness) is representative of the variation of brightness in the image.
This method was retrospectively applied to images from 708 subjects using a nested case-control design [22]. Relative risk estimates were moderate (typically ≥ 2.0 between the highest and lowest density categories) and the method was deemed successful, having the advantages of reproducibility and observer independence. However, the relative risks are still lower than those using interactive thresholding and visual assessment on the same study cohort [21].

Segmentation algorithms based on statistical methods should also be acknowledged, such as those developed by Karssemeijer [125], Heine and Velthuizen [126], Sivaramakrishna et al [127], Oliver et al [128] and Wang et al [129]. However, their widespread use is not reported.

One disadvantage to all techniques discussed so far is that they are only capable of quantifying the absolute or percentage area of dense tissue visible on a mammographic image. Although the density versus risk mechanism is not fully known, it is logical that risk is more closely associated with the actual amount of dense tissue within the breast rather than its projected area [32]; the mammogram is a two-dimensional projection of the breast and therefore only a representation of a three-dimensional structure. In fact, one author has gone so far as to state that area-based measures are scientifically invalid. Although these techniques are highly reproducible, this should not be confused with accuracy and segmented “dense” areas do not reflect the true percentages of tissue volume that created those densities [31]. Furthermore, these techniques cannot account for variations in compression force or x-ray exposure [130] and are highly dependent on breast positioning.

3.2 Volumetric analysis methods

Three methods are discussed in detail below: the model-based approach taken by Highnam and Brady [131 - 134], calibration techniques, such as the Manchester Method [36 – 38, 135] and a method utilising dual-energy X-rays [136, 137].

3.2.1 The “hint representation” and Standard Mammographic Form (SMF)

Highnam and Brady [131, 132] developed a unique image processing technique based on a model of image formation. In order to apply this technique to breast composition analysis, the breast is assumed to consist of fat and “interesting” tissue, such that the total breast thickness, \( H \), is defined as the sum of the thicknesses of the two components:

\[
H = h_{\text{int}} + h_{\text{fat}}
\]  \hspace{1cm} (3.1)

The intensity in any pixel \((x, y)\) on the digitised image indicates the amount of attenuation (absorption and scatter) in the column of tissue between the x-ray source and the pixel. After applying a scatter removal algorithm (details are beyond the scope of this thesis) [138], two
important equations are derived, representing the attenuation based on the intrinsic anatomy, $h \mu (E, x, y)$ and the energy imparted to the film from the primary beam, $E_p(x, y)$.

\[
h \mu (E, x, y) = h_{int}(x, y) \mu_{int}(E) + h_{fat}(x, y) \mu_{fat}(E)
\]
\[
= h_{int}(x, y)(\mu_{int}(E) - \mu_{fat}(E)) + H \mu_{fat}(E)
\]

(3.2)

\[
E_p(x, y) = \Phi(V_{\text{tube}}, x, y) A_p t_s \int_0^{V_{\text{tube}}} N_0^{\text{rel}}(E) S(E) G(E) e^{-\mu_{int}(E)h} e^{-\mu(E, x, y)} EdE
\]

(3.3)

where $\Phi$ is the photon flux at tube voltage, $V_{\text{tube}}$; $A_p$ is the pixel area; $t_s$ is the exposure time; $N_0^{\text{rel}}$ is the relative number of incident x-ray photons; $S(E)$ and $G(E)$ are the screen absorption ratio and grid transmission ratio for primary photons of energy $E$; $\mu$ is linear attenuation coefficient and $h$ is thickness. There are decreases in intensity caused by the Lucite compression plate and by the column of breast tissue above $(x, y)$. The linear attenuation coefficients used for fat and interesting (glandular) tissue are those of Johns and Yaffe [47].

Substituting Equation (3.2) into (3.3) yields only one unknown: $h_{int}(x, y)$. This is determined by equating the primary energy found in the practical case with the theoretical value and solving the resulting non-linear equation. Some calibration data are also required to achieve this. A Lucite stepwedge is exposed on a film with a Lucite block over the Automatic Exposure Control (AEC) chamber; this enables determination of the relationship between the film density and $E_p$. This film is also digitised so that pixel value can be related to film optical density.

In order to generate the “$h_{int}$ representation”, it is necessary to know the kVp, mAs and compressed breast thickness, H for every image. Knowledge of H is particularly crucial in accurately determining the volume of dense breast tissue. The method employed [139] is summarised as follows, with reference to Figure 3.4, which depicts the CC examination (left) and the resulting mammogram (right).

It is assumed that the breast periphery consists entirely of fat. Arc A lies on the outer edge of the breast and represents the point at which optical densities in the film background (i.e. attenuation by air) start to become lighter. At arc B, film optical density is lower than that at arc A, as the breast thickness and hence attenuation increases. However, the tissue composition is still homogeneous (entirely adipose) so the iso-intensity curve of optical density would be smooth. Moving towards the chest wall, these iso-intensity curves would reflect the decreasing film optical density but would remain smooth up to and including arc C, which represents the inner edge of the breast margin. Beyond this point the iso-intensity curves are non-smooth because they represent the optical densities within heterogeneous tissue (e.g. arc D). Arc C is used to
determine the breast thickness, $H$, based on the assumption that $h_{\text{int}} = 0$ and therefore $h_{\text{fat}} = H$. Note that a correction is made to account for the divergence of the x-ray beam, since the thickness of tissue between the x-ray source and point C in Figure 3.4 (left) is not the same as that denoted by $H$.

This technique was compared with breast thickness measured by hand, using a ruler, during mammography examinations. Measurements were felt to be accurate to within ±0.3cm for the CC view and ±0.5cm for the MLO view. The average absolute difference between the measured and estimated value was 0.22cm (maximum 0.71cm) for the CC view and 0.44cm for the MLO view [139].

By 2006, the authors had made several advances and the technique became known as the Standard Mammographic Form (SMF) [134]. The most notable development was “Calibration Parameter Compensation” (CPC) which enabled the technique to be applied retrospectively. The method of breast thickness determination was actually used as the basis for CPC. Starting with the initial assumption that $H = 1.66 + (0.034 \times \text{breast area})$, iso-intensity curves were assigned within the breast edge and the point at which they stopped being smooth ($h_{\text{int}} = 0$) was used as the final estimate of $H$. This point is considered “ground truth”. In the absence of calibration data, a number of other assumptions were made, including a beam quality of 28kV, Mo/Mo, a film optical density of 1.6 and the breast tissue over the AEC chamber being composed of 30% interesting tissue. The outputs are $\text{SMF}_{\text{vol}}$, the absolute volume of interesting tissue within the breast, and $\text{SMF}_{\%}$, the volume of interesting tissue expressed as a percentage of the breast volume.

SMF was used on the mammograms of 626 women (3,816 images) [130] and demonstrated a strong positive relationship between the Boyd SCC (visual assessment) [14] and $\text{SMF}_{\%}$ but not
SMF$_{\text{vol}}$. SMF$_{\text{vol}}$ was found to be significantly lower in the CC view compared with the MLO view ($p < 0.001$) but SMF$_{\%}$ was higher in the CC than the MLO view, although these differences were not significant ($p = 0.69$). Agreement between the left and right breasts was good, with Pearson correlation $r = 0.92$ for $\ln$(SMF$_{\text{vol}}$) and $r = 0.85$ for $\ln$(SMF$_{\%}$). A later study on 250 women (1,000 images) revealed no significant difference in the average SMF$_{\%}$ of left and right breasts, but for individual women, the correlation between the left and right breasts was 0.77 in the CC view and only 0.68 in the MLO view, compared to 0.90 using interactive thresholding [140].

In a study on the MLO images of 590 women [141], SMF was found to exhibit a relationship with a small number of breast cancer risk factors. Both SMF$_{\text{vol}}$ and SMF$_{\%}$ were found to be positively and significantly correlated with age at last menstrual period. Interestingly SMF$_{\text{vol}}$ was positively correlated with BMI whereas absolute area and percentage density by area and volume exhibit the opposite effect [81, 141]. SMF$_{\text{vol}}$ was also weakly associated with family history and showed significant negative relationships with parity and number of births, in line with the fact that increasing parity is associated with decreased breast cancer risk [79]. These findings suggest that absolute glandular volume may be an important parameter, in addition to percentage volumetric density.

The advantages of this method are the sophistication of the model of image formation, which even includes scatter removal, and the ability to be applied retrospectively; calibration techniques can only be used prospectively. However, the authors acknowledge that having calibration data generally provides more accurate results [134]. Given that SMF is fully automated, it is 100% repeatable but not 100% accurate in its breast segmentations. In one study, SMF failed on 24 out of 1,000 images, with the failure rate rising with increasing density [130]. A further advantage is that the algorithms developed can be applied to digital mammography, as long as the raw (“for processing”) images are available. This has already been achieved and the methods are described in Section 3.3.1.1.

The disadvantages are the relatively poor correlation between left and right breasts for individual women [140]. This suggests that an average SMF measure of the left and right breasts, or even of all four views, should be used. Another potential concern is the accuracy of breast thickness estimation, given that a 0.1cm error in $H$ could lead to a 5% error in volumetric percentage breast density [134].

SMF has exhibited a strong relationship with the Boyd SCC [14] using both visual assessment and thresholding techniques [130, 140] and the Wolfe scheme [134] and SMF$_{\%}$ has shown a strong negative correlation with BMI [141]. However, it has not demonstrated an association with other risk factors such as age (after adjustment for BMI) or current HRT use [130] and disappointingly, its relationship with breast cancer risk was found to be weak and non-significant [33]. In a case-control study with 634 cancers (206 intervals and 428 screen-detected) and 1,880 age-matched
controls, an Odds Ratio (OR) of 1.92 was found between the highest and lowest quartile of SMF%, compared to 2.45 for PMD using thresholding. However, when considering only the screen-detected cancers, the OR for SMF% had dropped to 1.23 and failed to reach significance (p = 0.194); the OR for PMD was 1.56 and retained significance (p = 0.007). A possible reason as to why SMF is not strongly correlated with breast cancer risk is that the method tends to fail in women with greater breast density which may introduce biases as these women are at higher risk [130].

### 3.2.2 Calibration techniques

The Manchester Method [36 – 38, 135] is a semi-automated technique which involves imaging a calibrated stepwedge alongside the breast for every mammogram. A full description is given in Chapter 4, but the main points are summarised below.

Calibration is achieved by acquiring x-ray images of the stepwedge imaged alongside phantoms constructed of adipose (AP6) and glandular (WT1) tissue-equivalent material [142] encompassing a range of total breast thicknesses and compositions. This removes any potential errors associated with published values of the attenuation coefficient of the stepwedge material, which will be energy-dependent. Following digitisation of images, the grey level within the phantom is matched to the stepwedge thickness with the same grey level. A calibration film using the original PTFE (Teflon) stepwedge is shown in Figure 3.5.

![Figure 3.5: Example calibration film from the original Manchester Method [36]](image)

The calibration data are used to generate a surface relating stepwedge thickness, glandular tissue thickness and total breast thicknesses. In order to measure total breast thickness, four pairs of radio-opaque markers are placed on the compression paddle. Using simple magnification geometry, the distance between the markers on the image is used to calculate the breast thickness at each of the four points and a plot of these determines the smooth decrease in breast thickness.
thickness across the compressed breast region. This has the advantage of accounting for paddle tilt, which has been shown to be non-trivial [143, 144]. The breast edge has a rapidly decreasing thickness and this is modelled as an ellipse. All clinical images must be acquired with both the stepwedge and markers and as such, the technique can only be used prospectively. Digitised images are analysed using software written in Matlab (The MathWorks, Inc) which segments the breast from the background and then matches the grey level of every pixel in the breast region to the thickness of stepwedge material with the same grey level. Combined with the measured value of total breast thickness, calibration data enables glandular tissue thickness to be uniquely determined at each pixel. This is summed over the breast region to give the absolute volume of glandular tissue. This is also divided by breast volume to generate a percentage measure of volumetric breast density.

A similar technique has been developed by Pawluczyk et al [145] although there are some notable differences in the calibration procedure and the breast thickness measurement. The calibration device is a complex stepwedge which has steps of height 1 – 8cm in 1cm increments and contains 5 different compositions of tissue-equivalent material: 0%, 30%, 50%, 70% and 100% glandularity. For each calibration image, 40 different grey levels are displayed. The first step in the calibration procedure is to determine clinically relevant exposure factors (kV, target, filter and mAs). This is achieved by imaging slabs of tissue-equivalent plastics [146] under Automatic Exposure Control (AEC).

These derived exposure factors are then set manually with the calibration device positioned in the usual breast location and a small aluminium stepwedge (heights 1.5 – 10.5mm in 1.5mm increments) positioned in the corner of the image. For each kV / target / filter combination, at least three mAs values were chosen. These images were digitised and calibration surfaces were generated to show the relationship between percentage gland, total breast thickness and the log relative x-ray absorption (LRXA).

Ideally, the calibration device would be imaged alongside the breast for every image but there are obvious physical limitations, both in height and area. Therefore, only the aluminium stepwedge is included alongside the breast and this informs changes from the initial calibration conditions, so that the surface data may be shifted as appropriate. Knowledge of the exposure factors for each image is required.

The system-indicated breast thickness is used but constraints are applied to this value based on LRXA. The minimum value of LRXA would be expected to correspond to the path through the maximum thickness of dense tissue and the maximum LRXA would correspond to the path through the least amount of dense tissue. If the minimum and maximum values of LRXA do not lie within LRXA_{100\%\text{gland}} and LRXA_{100\%\text{fat}}, then the breast thickness would be adjusted so that these conditions were satisfied. No account was made for paddle tilt.
The authors state that phantom studies have shown volumetric breast density to be accurate to within 5% of the true value when using the actual calibration curves. However, care must be taken in interpreting percentage errors when the unit of measurement is also “%”. The accuracy of the technique is excellent at 0% density and ≥75% density but less so at intermediate densities. For example, using a 2cm thick phantom with 25% density, the actual calibration curves gave a result of 20% and the shifted calibration curves gave a result of 18%, which equates to a percentage error of 28%.

This method is now referred to as Cumulus V (Volumetric). It was used in a prospective case-control study with 364 screen-detected cases and 656 controls [34] where its predictive power for breast cancer risk was compared to interactive thresholding [60] using Cumulus v4. A number of observations were made, the first being that there were more cases than controls in the highest quintile of breast density compared to the lowest, and more controls than cases in the lowest quintile than the highest. This was true for absolute and percentage measures by area and volume. Furthermore, tests for the trend of increasing risk with increasing percentage density were strongly statistically significant. After adjustment for risk factors such as age and BMI, ORs for the 5th quintile compared with the 1st quintile were 1.98 and 1.86 for percentage density by volume and area respectively. Both measures were significantly associated with risk when treated separately but when both were included in a predictive model, only the area-based measure retained significance.

Note that the technique described by Pawluczyk [145] for breast thickness estimation was improved prior to application in the case-control study, given the inaccuracies associated with the indicated breast thickness readout and the finding that an error in breast thickness measurement of 3mm would result in an error in percentage volumetric density of at least 15% [147]. A technique known as optical stereoscopic photogrammetry (OSP) overcomes these limitations [148]. Two small webcams are mounted to the tube housing and a thin, optically transparent plastic film, with a grid of dots (radiolucent but with high visual contrast) is mounted on the compression paddle. A checkerboard pattern is also placed on the breast support platform during calibration (but not during the acquisition of mammograms) in order to measure and inform corrections for camera geometry and distortions. The method employs a similar principle to stereotactic imaging in mammography; that is, two visible light images are acquired of the top surface of the compression paddle from known fixed locations. The parallax shift of the positions of the dots in each of the stereo-views is used to determine the separation of the “dot” from the plane of the breast support plate. These images are acquired at the moment of x-ray exposure. Only dots in the area of contact between the breast and the compression paddle are included. OSP has been found to be highly accurate (mean error ≤ 1mm) and is independent of applied compression force and the breast positioning [148]. Unfortunately, it is labour intensive, highly dependent on room lighting, requires informed consent and cannot be used retrospectively. It is
therefore considered an excellent research tool and has been used as the “gold standard” to guide the development of a practical secondary approach.

The secondary approach [149] was employed in the case-control study [34] and uses system-indicated thickness and OSP-measured thickness of phantoms to generate prediction equations for the estimation of local breast thickness, using system-indicated breast thickness as the initial input. Referring to Figure 3.6 below, the prediction equation (3.4) is:

\[ Z_p = Z_r + f(x, y, F_r) \]  

where \( Z_p \) is predicted thickness, \( Z_r \) is machine-reported thickness and \( f(x, y, F_r) \) expresses tilt and bulge of the lower surface of the paddle, both of which are dependent on force.

This method has been shown to be accurate to within 5mm, which may be better than the system-indicated value, but could still result in significant errors in percentage breast density [147].

The main advantage of calibration techniques is that volumetric breast density quantification is based entirely on measured data rather than a dependence on models, thereby accounting for inter-image changes in exposure factors. Additionally, breast thickness measurement is considered very accurate and accounts for paddle deformation. However, the fact that these techniques can only be applied prospectively is clearly a disadvantage. Furthermore, the test objects may not always fit alongside the breast. Currently, the Manchester Method is semi-automated, and although the Pawluczyk method [145] is said to be fully automated, it does require knowledge of exposure factors which, if available, are stamped on the film, so some user input is needed to link these factors to the calibration data. Digital mammography may offer solutions to both of these issues.

Unfortunately, as with SMF [33], the Pawluczyk method [145] has not shown a strong and statistically significant relationship with breast cancer risk [34]. To date, the Manchester Method
has only been applied in a small-scale study (39 participants) investigating the relationship between weight change and density [150].

3.2.3. Dual energy X-ray techniques

A feasibility study on the use of Dual Energy X-ray Absorptiometry (DXA), routinely used for the assessment of bone density and whole body composition, was carried out by Shepherd et al [136, 137] with the hypothesis that DXA would provide a more accurate and precise measurement than the thresholding techniques employed in mammography. In addition, the radiation dose is extremely low (entrance dose 0.3mGy), breast compression is not required and no subjective interpretation is necessary. The aim of the study was to calibrate a commercially available DXA scanner to measure breast glandular density on phantoms and in mastectomy samples and compare the results to conventional mammographic density measurements. The parameter “1 - % fat” as measured by DXA was plotted against the known percentage of glandular tissue within the breast phantom and was found to be highly correlated and linear. However, the gradient was not equal to one, the reason being that the parameter “% fat” was measured relative to a two-component model of fat and muscle as opposed to fat and glandular tissue. Precision was thought to be 1% and was found to be limited by re-positioning rather than noise. Although results were relatively promising, the calibration of the scanners is not entirely suitable for routine clinical use and DXA images have no additional diagnostic value.

Dual energy mammography has been developed by several equipment manufacturers and may prove to be a valuable tool for breast composition analysis in the near future.

3.3 Challenges of digital mammography

Film-screen mammography will soon be completely replaced by Full Field Digital Mammography (FFDM) in the NHSBSP. It is therefore important to consider whether the techniques described in Sections 3.1 and 3.2 will remain valid. The appearance of digital mammograms is noticeably different to that of film-screen mammograms and this has implications for breast density assessment using area-based methods (Section 3.3.2). Two images are available based on a single x-ray acquisition. The first is the raw (“for processing”) image; the second is the processed (“for presentation”) image which is used by the radiologists for diagnosis.

For equipment with an integrated digital detector, the raw image has pixel intensities which are proportional to the x-ray transmission through the breast. A plot of mean pixel value against detector dose would provide a linear response, rather than the S-shaped characteristic curve associated with film. This is certainly advantageous for calibration techniques because, as shown by Kaufhold et al [151], it removes the need for a stepwedge. There are a number of further advantages for volumetric techniques as discussed in Section 3.3.1; in particular, there is no
requirement to digitise films, making the analysis even faster. The only disadvantage associated with using raw images is that they are not routinely sent to the Picture Archiving and Communications System (PACS) due to cost and storage limitations.

3.3.1 Volumetric techniques for FFDM

The advantages offered by FFDM for volumetric density analysis include the following. Firstly, calibration data are always available in the DICOM (Digital Imaging in Communications and Medicine) header [152]. Secondly, there is no need to apply a separate correction algorithm for the anode-heel and inverse square law effects as even the raw images have had a “flat-field” correction applied, ensuring uniform pixel intensity across the image. Finally, image quality improvements offered by digital mammography should enable more reliable segmentation of the breast outline, the inner breast edge and the pectoral muscle. A brief discussion is presented below with reference to model-based methods and calibration techniques.

3.3.1.1 Quantra™ and Volpara™

These methods are extensions of SMF [134]. Quantra™ [153] has been found to offer improvements in estimating volumetric breast density in dense breasts. SMF tended to fail in these women as finding an area of 100% fat, which was the basis for determining breast thickness, proved to be difficult in these women. The result was that they were excluded from clinical studies, undermining the predictive power of SMF. The correlation between breast density in left/right and CC/MLO views is also better using Quantra™. Quantra™ is commercially available and as such, literature on the algorithms employed is unavailable. It is therefore unclear whether these improvements are the result of more sophisticated modelling, or simply the better visualisation of dense breast tissue offered by the higher energy x-rays and increased dynamic range of digital mammography.

Similarly, Volpara™ [154] is also commercially available so details are limited. It employs relative, rather than absolute, physics modelling and one of the main changes is in the use of the method previously described to estimate breast thickness [134, 139]. The pixel value corresponding to the area of entirely fatty tissue \( P_{\text{FA}} \) is now used as a reference level to find the thickness of dense tissue \( h_{d} \) at each pixel using equation (3.5). It can be assumed that for digital detectors, pixel value \( P \) is linearly related to the energy imparted to the x-ray detector.

\[
\ln\left(\frac{P(x, y)}{P_{\text{FA}}}\right) = \frac{\mu_{\text{fat}} - \mu_{\text{dense}}}{\mu_{\text{fat}} - \mu_{\text{dense}}} h_{d}(x, y)
\]

Breast thickness is taken from the indicated readout, with a fixed slant assumed for the compression paddle [155]. The authors acknowledge the inaccuracies associated with indicated readout, but because the error would affect breast volume and dense volume in the same way in
the current implementation, percentage breast density would not vary widely with changes in breast thickness. Additionally, the indicated readout of breast thickness tends to be underestimated and therefore breast density will be overestimated, meaning that the woman would never be treated in a lower risk category, although this could result in unnecessary intervention. Volpara™ has shown excellent accuracy when assessing phantom densities [154] and studies on its association with breast cancer risk are underway. Volumetric breast density measurements are lower using Volpara™ than SMF and Quantra™, as the skin is excluded from the density measure.

3.3.1.2 Calibration techniques

A method has been developed by Kaufhold et al [151] which shares many features with those described in Section 3.2.2 with the notable difference that the use of a tool such as a stepwedge appears to be unnecessary and the digital detector itself is used as the calibration device. The generation of calibration curves involves exposing tissue equivalent material under a wide range of manually selected clinical exposure parameters, to cover a range of breast thicknesses and compositions (0%, 50% and 100% glandularity). Although the calibration initially provides information in terms of pixel intensity (grey level) in the raw image, it is necessary to convert these values to the negative log domain due to the exponential attenuation relationship between detected photons and the material thickness. The values are also normalised to a specified mAs (110 mAs was chosen).

The authors only consider the compressed breast region, which is assumed to have a constant thickness of H cm throughout. The method was applied to 23 mammograms and a theoretical error analysis showed that the greatest source of error was inaccuracies in the estimate of compressed breast thickness, with a 2mm inaccuracy resulting in a 7% error in percentage breast density measurements.

Such a technique has the advantage of being applied retrospectively. However, further investigation is necessary to confirm that detector performance remains stable over extended periods of time. Additionally, there must be a suitable method of monitoring stability, in order to determine when recalibration is required. Kerrison et al [156] found that the mean pixel value (MPV) and mAs values, measured during daily quality control tests, could be used for this purpose. Detector performance was found to be remarkably stable over long periods of time (exceeding 244 days) and substantial deviations from the baseline value, which tended to coincide with service engineer visits, could be used to inform the need for recalibration.

Because of concerns associated with detector stability, Malkov et al [157] believe that including a calibration device in the images remains preferable. Their technique is known as Single X-ray Absorptiometry (SXA) and uses a CIRS breast tissue-equivalent phantom [146] with a glandularity
of 80%. The phantom has steps of height 1 – 7cm and has nine lead markers around the top, middle and bottom (Figure 3.7a). The markers enable accurate measurement of breast thickness and paddle tilt. The SXA phantom is positioned on the compression paddle in the unused portion of the mammogram (Figure 3.7b), and would overly the breast in <1% of images.

![Figure 3.7: (a) SXA phantom; (b) the phantom as seen on the digital image [157]](image)
Reproduced with permission from Medical Physics, AAPM

As with other calibration techniques, the reference phantom is calibrated against tissue equivalent material and the outputs are absolute fibroglandular volume, FGV, and percentage fibroglandular volume (%FGV). In a similar manner to Kaufhold et al [151], attenuation images are calculated by a log formula and normalised to 100mAs.

Using specially designed phantoms, Malkov et al [157] found that the SXA phantom had a breast thickness accuracy of 0.4mm and a tilt accuracy of 0.09°. The system-indicated thickness for phantoms was on average 9.7mm greater than the SXA measure. The error in %FGV was up to 7% per 1mm thickness deviation and increased with decreasing breast thickness. Interestingly, although breast thickness is commonly quoted as the greatest source of error in breast density measurement, the tilt appears to have a far more significant effect. Failure to account for paddle tilt can give a 40% error in %FGV (2cm breast, 6° tilt) but an error of only 2% when using the SXA phantom [157]. The technique has shown good left / right breast agreement with r = 0.89 for %FGV and r = 0.77 for FGV and correlates well with percentage mammographic density measured using thresholding software developed in-house.

Cumulus V has also been adapted for digital mammography [158, 159] and has been applied to the simulated 2D mammograms of 26 women. These were based on the 3D dataset acquired on a prototype dedicated breast computed tomography (DBCT) scanner. Within each slice, the tissue was segmented into air, skin, adipose tissue and fibroglandular tissue. The mammograms were simulated with and without skin inclusion. The mean absolute difference between DBCT and Cumulus V measurements on mammograms including skin was 2.1% (95% CI: 1.3 – 3.0).
3.3.2 Area-based density assessment using FFDM

Radiologists use the processed (“for presentation”) image for diagnosis and report that breasts tend to appear less dense on these images than on film [115]. This is certainly advantageous for cancer detection; the ACRIN-DMIST trial [160] found that digital mammography was more sensitive in women with dense breasts, including those under 50 and those who were pre- or peri-menopausal. However, this presents problems for density assessment, especially if women are taking part in longitudinal studies and prior density measurement has been made on film. Figure 3.8 shows the same breast imaged using film and digital imaging.

![Figures 3.8: Digitised film priors (left) and processed digital mammography images (right)](image)

The film priors were digitised using the Array MammoPro. Processed digital mammography images appear less dense, particularly the CC view which also illustrates how changes in positioning affect percentage mammographic density; this poor reproducibility is one of the criticisms of using area-based techniques. These images were taken one year apart so the decrease in density cannot be completely attributed to age.
Results for mammographic density measurement using Cumulus v4 are also different for digital mammography compared to film. Jeffreys et al [119] compared the results of Cumulus v4 on digitised film to Cumulus v4 on the raw digital image. One radiologist evaluated the LCC view of 324 asymptomatic women from the screening population. All women had analogue and digital mammography images acquired on the same day. Percentage mammographic density on the raw digital image was found to be systematically over-estimated compared with film image. A 1% increase in density in the digital image was equivalent to a 0.89% increase in the film image and the mean difference between measures was 3.96%; the correlation coefficient between the two measures was 0.95.

Interestingly, the radiologist who evaluated these images came to the opposite conclusion when using processed digital images. Harvey [161] compared the results of Cumulus v4 on 60 digital images with standard post-processing to the results for the 60 prior film images which had been acquired within the past 2 years. The LCC view was used and all cases were diagnosed as normal. This time, the percentage mammographic density was found to be significantly lower for digital mammograms (mean density 32.2%) compared to analogue (mean density 40.3%). An additional finding was that the difference increased with density. For women with 10.1 – 20% mammographic density, the average difference was only 2.7% between film and digital, increasing to 14.9% for women with 60.1 – 70% density. Once again, there was excellent correlation between the two methods, with a linear regression strength of fit, $r = 0.96$.

Although a slight decrease in density with age would be expected, the main reason for the difference is attributed to the fact that the skin line is better identified on digital mammograms resulting in the inclusion of more subcutaneous fat and therefore lower percent density. However, in some cases, the skin edge is so well defined that it appears as a bright white line, as shown in Figure 3.9 and this could be mistakenly classed as dense tissue when using a thresholding method.

![Figure 3.9: The boundary between the compressed dense breast region and the layer of subcutaneous fat in the breast edge is distinct and the skin edge appears as a bright white line](image_url)
One concern is that if breast density is underestimated when using digital mammograms compared to analogue, then breast cancer risk may also be underestimated. However, the measures of breast density generated by Cumulus v4 on processed digital images have been found to be related to clinical experience. In a study by Conant et al [117] three breast radiologists and three medical physicists trained in the use of Cumulus v4 each analysed 40 processed digital mammography images. These were from women with previously diagnosed breast cancer but the contralateral MLO view was used for analysis. Inter-reader agreement, assessed using the Pearson correlation coefficient, was high among both the clinical group (r = 0.91, p < 0.001) and the non-clinical group (r = 0.83, p < 0.001). However, the non-clinical group consistently measured higher mammographic density (mean = 35%) than the clinical group (mean = 24%), which was statistically significant. This is contrary to the results of Byng et al [60] who found excellent inter-observer agreement between radiologists, clinical scientists and the ‘naïve observer’. Further work is required to assess whether this is a genuine effect of digital mammography or if the result is isolated to this study. The worst case is shown below in Figure 3.10. A difference of this magnitude could impact on a patient’s risk assessment outcome, although this is only one group difference using the Boyd SCC [14].

![Figure 3.10: Left to right: processed mammogram; Cumulus used by radiologist; Cumulus used by physicist [117]. Reproduced with permission from Springer, RightsLink Copyright Clearance Center](image)

It is interesting to note that one of the developers of Cumulus recommends that the raw images are processed using specially developed algorithms which emulate film-screen characteristics [115]. This further increases the burden of human resources on a labour-intensive technique.
3.4 Discussion

Inevitably, there are advantages and limitations associated with both area-based and volumetric techniques. Although area-based methods are commonly criticised for their subjectivity, inter-reader reliability tends to be high with reader training. Percentage mammographic density assessed visually and using interactive thresholding consistently shows a strong relationship with breast cancer risk and interactive thresholding is frequently described as the “gold standard” for this reason. However, there are concerns that the strength of this relationship will be lower for digital mammography.

Volumetric methods are objective and translate well to digital mammography. As long as the raw images are available for analysis, absolute and percentage glandular volume are available at the touch of a button. Unfortunately, their association with breast cancer risk is not well-established. Only small clinical studies have been carried out to date and the strength of association was found to be far lower than that of interactive thresholding. One reason may be that the methods are prone to fail in women with the densest breasts and excluding these women will inevitably introduce bias. However, digital mammography should resolve this issue and algorithms will be improved with the addition of data extracted from the DICOM header. Volumetric techniques have shown strong correlations with thresholding [118, 119] and it is anticipated that the strength of association between volumetric density and breast cancer risk will be established, as it is considered to be a more scientifically valid parameter than mammographic density. The success of these techniques is highly dependent on accurate measurement of breast thickness and paddle tilt.

Truly volumetric modalities such as MRI and DBCT may be the ideal choice for accurate glandular volume measurement. However, these are typically diagnostic imaging tools, making it difficult to collect sufficient data to determine the relationship between density and risk. Advanced mammography techniques such as tomosynthesis and dual energy imaging may be beneficial but are not currently in widespread use.

Chapter 4 describes the Manchester Method in detail, with particular emphasis on the extensions carried out for this PhD, most notably the design of a new stepwedge.
4. Development and Extension of the Manchester Method

The original Manchester Method was developed by Smith et al [36, 162], with adaptations made by Marchant [37] and Patel [163]. Previously, the method had only been used in a small research study with 39 participants, which evaluated the effect of diet and exercise on breast volume and density [150, 163]. Improvements achieved during this PhD have facilitated the application of the method in the breast screening environment.

In this chapter, the underlying theory is presented, followed by a detailed discussion of each aspect of the methodology: measurement of compressed breast thickness, design of a suitable calibration device, collection of calibration data, breast segmentation and modelling of tissue thickness in the breast margin. A description of the current analysis software is given, including screen-shots. Finally, the advantages and limitations of the method are discussed.

4.1 Theoretical principles

The method requires a calibration device (stepwedge) to be imaged alongside the breast during the x-ray examination. Following digitisation of the mammograms, the grey level of every pixel within the breast is matched to the same grey level within the stepwedge. The equivalent thickness of stepwedge material, combined with knowledge of total breast thickness, enables breast composition to be uniquely determined at each pixel.

It is assumed that the breast is composed of only two types of tissue: adipose (fatty) and dense (glandular). For the column of tissue above any pixel in the mammogram, the total thickness of breast tissue $x_b$ is composed of thicknesses $x_g$ of glandular tissue and $x_f$ of fatty tissue:

$$x_b = x_g + x_f \quad (4.1)$$

The intensity ($I$) of the x-ray beam after passing through a thickness $x_b$ of breast tissue is given by equation (4.2):

$$I = I_0 e^{-\left(\mu_g x_g + \mu_f x_f\right)} \quad (4.2)$$

where $\mu_g$ and $\mu_f$ are the linear attenuation coefficients of glandular and fatty tissue respectively, and $I_0$ is the intensity of the incident x-ray beam.

The intensity of radiation after passing through a thickness $x_{sw}$ of stepwedge material with attenuation coefficient $\mu_{sw}$ is:
\[ I = I_0 e^{-\mu_{\text{sw}} x_{\text{sw}}} \quad \text{(4.3)} \]

The thickness of stepwedge material causing the same amount of attenuation as breast thickness \( x_b \) can be determined by equating (4.2) and (4.3):

\[ \mu_{\text{sw}} x_{\text{sw}} = \mu_g x_g + \mu_f x_f \quad \text{(4.4)} \]

Substituting equation (4.1) into (4.4) yields an expression for \( x_g \) in terms of only known quantities.

\[ x_g = \frac{\mu_{\text{sw}} x_{\text{sw}} - \mu_f x_h}{\mu_g - \mu_f} \quad \text{(4.5)} \]

Strictly speaking, these equations are only true for mono-energetic x-rays but the mammography x-ray spectrum is poly-energetic. However, it does have a narrow beam of energies due to the use of k-edge filters (Figure 2.4).

The accuracy of published values for the attenuation coefficient of the stepwedge material may not be guaranteed for all materials. Furthermore, they are energy dependent. Therefore, calibration data are generated from the x-ray images of the stepwedge imaged alongside tissue-equivalent phantoms covering a range of thicknesses and compositions. This process is described in Section 4.4.

The current method does not include corrections for the effects of scattered radiation or the anode-heel effect. The implications of this are discussed in Chapter 5 (Uncertainties in the Manchester Method).

### 4.2 Measurement of compressed breast thickness

All mammography units provide an electronic reading of indicated breast thickness and there is a UK quality assurance requirement that this is accurate to within ±5mm [164]. The automatic selection of optimum exposure factors, particularly tube voltage (kV), is often based on this value and it is also used in the calculation of mean glandular dose. However, as discussed in Chapter 3, an accuracy of ±5mm introduces unacceptable errors into volumetric breast density measurement [144, 147, 148, 157]. Additionally, the indicated breast thickness is highly dependent upon compression force and has been found to vary by up to 15mm on some units when compressing the same breast or phantom [148]. Furthermore, a single reading of breast thickness cannot account for paddle tilt, which has also been shown to lead to significant inaccuracies in measuring volumetric breast density [157].
The method employed for measuring compressed breast thickness is based on the work of Burch and Law [165]. Four pairs of radiopaque (lead) magnification markers are glued to an acetate sheet that fits snugly on the compression paddle. Using the magnification geometry shown in Figure 4.1, the separation between the markers on the film, $x_f$, can be used to determine the compressed breast thickness, $x_b$, at a number of points, thereby accounting for paddle tilt.

![Figure 4.1: Geometry used to determine breast thickness, using radio-opaque markers placed on the compression paddle](image)

The separation of lead discs placed on the compression paddle will be magnified on the image by an amount $M$, given by:

$$M = \frac{x_f}{x_p} = \frac{FID}{FID - x_b}$$  \hspace{1cm} (4.6)

Rearranging equation (4.6) yields an expression for breast thickness, $x_b$.

$$x_b = FID - \frac{FID}{x_f} x_p$$  \hspace{1cm} (4.7)

It can be difficult to measure the focus to image distance ($FID$) as it is not possible to know the exact position of the focal spot (although the approximate location is marked by a dot on the tube housing) or where the film lies in the cassette. Furthermore, the method described above would result in a breast thickness estimate which includes the paddle thickness and the distance from the top of the breast support platform to the film. Errors associated with these inaccuracies are prevented by calibrating marker sheets against known thicknesses of rigid material, as described in Section 4.2.2.
4.2.1 Construction of new marker sheets

Marker sheets consist of four pairs of lead discs of diameter 3mm and depth 2mm glued to a piece of acetate (normally used on overhead projectors). The sheets from previous studies were no longer suitable for use as they had been designed for a Lorad mammography unit and had since become damaged. Two GE (General Electric) mammography units would be used in the feasibility study, each with an 18x24cm and 24x30cm compression paddle. It was necessary to cut the acetate to a size which fitted tightly inside the compression paddle (but without bending), as the mediolateral oblique (MLO) views are acquired at an angle of 45 - 50° and the marker sheet could slide around or fall out.

It was also necessary to review the location of the lead markers as they were not visible on films where breast thickness was $\geq$ 80mm, hence new sheets were designed for this PhD. Sections 4.2.1 – 4.2.3 represent original work.

The markers must be positioned so that the possibility of being obscured by an ID label, view marker, cassette label or stepwedge is minimised. Figure 4.2 shows the typical film layout. More importantly, the markers should not overly breast tissue. Unfortunately, it is impossible to avoid this altogether because if the markers are placed too close to the edge of the compression paddle, the magnification geometry associated with increasing breast thickness means that they will not be present on the image. After reviewing a number of mammograms, the marker locations were chosen to be 4, 8, 12 and 16cm back from the chest wall at a distance of 2.3cm and 3.0cm from the edge of the compression paddle on 18x24cm and 24x30cm field sizes respectively. In theory, this would enable compressed breast thickness to be measured up to 9.4cm on 18x24cm films and 10.4cm on 24x30cm films, so very few women would be excluded.

![Figure 4.2: Typical marker positions. The possibility of markers overlying breast tissue, or being obscured by other objects on the image must be minimised](image-url)
The feasibility study was carried out at two sites. Rather than having to write different analysis software for each site, it was hoped that a single set of thickness calibration data could be used for all films acquired during the study. It was therefore necessary to construct sheets which were identical.

As a first attempt, the acetates were overlaid on graph paper with the marker locations indicated. The lead discs were glued to the acetate at each location using tweezers to position the discs. Unfortunately the sheets were sufficiently different that an appreciable error would be introduced in the measurement of breast thickness if one set of calibration data were used, based on the average of both sets. The error in measurement was found to be a maximum of 1.3mm for the 18x24cm sheets and 3mm for the 24x30cm sheets.

The solution was a jig constructed by staff in the mechanical workshop at the Christie Hospital, Manchester. This was a sheet of Perspex with holes drilled in it at the locations required for the lead discs. This could be overlaid on the acetate and the lead discs dropped into the holes, thereby ensuring that they would lie in an identical location on each sheet.

4.2.2 Calibration of magnification markers

Marker sheets were placed on the compression paddle and known thicknesses of tissue equivalent material were compressed very lightly so that the compression paddle just rested on the phantoms. The compression paddle is normally controlled electronically using a foot pedal but very fine adjustments in compression height and force can be made using a manual control which was therefore used. The thickness of the material was measured to the nearest 0.1mm using vernier callipers. The calibration films were digitised at a pixel size of 44μm and the separation between the markers was measured as the distance from the centre of one marker to the centre of the other marker within that pair. The relationship between marker separation in pixels and breast thickness in mm is shown in Figure 4.3 for both field sizes (18x24cm and 24×30cm) at the Bury site.

Pair 1 is closest to the nipple edge; pair 4 is closest to the chest wall. The phantoms were made of rigid material which is why the data for each marker pair are so close. On real breasts, the separation of pair 4 (chest wall) would be expected to be greater than that of pair 1 (nipple edge) as the compression paddle would tilt due to the decreasing breast thickness in the chest wall to nipple direction.

A linear trendline (y = mx + c) was fitted to the data from each marker pair for each sheet. Calibration data for pair 2 are shown in Figure 4.4 for the 18x24cm field size at Bury and Bolton.
Figure 4.3: Marker separation (in pixels) versus phantom thickness

Figure 4.4: Calibration data for marker pair 2 on 18x24cm field size
Pink: Bury site; Blue: Bolton site
All calibration data are shown in Table 4.1. An average of \( m \) (gradient) and \( c \) (intercept) was taken to produce a single set of calibration coefficients which were incorporated into the Matlab programme for breast density calculation. Unfortunately, one marker from pair 1 was rarely present on the Bury 18x24 images as it could not be seen in the film ID label and the decision was made to only use the three pairs of markers closest to the chest wall for breast thickness measurement on this field size.

<table>
<thead>
<tr>
<th></th>
<th>18x24</th>
<th>18x24</th>
<th>18x24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bury</td>
<td>Bolton</td>
<td>Average</td>
</tr>
<tr>
<td></td>
<td>( m )</td>
<td>( c )</td>
<td>( m )</td>
</tr>
<tr>
<td>Pair 1</td>
<td>8.099</td>
<td>4475</td>
<td>8.099</td>
</tr>
<tr>
<td>Pair 2</td>
<td>8.300</td>
<td>4475</td>
<td>8.399</td>
</tr>
<tr>
<td>Pair 3</td>
<td>8.325</td>
<td>4480</td>
<td>8.383</td>
</tr>
<tr>
<td>Pair 4</td>
<td>8.346</td>
<td>4483</td>
<td>8.370</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>24x30</th>
<th>24x30</th>
<th>24x30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bury</td>
<td>Bolton</td>
<td>Average</td>
</tr>
<tr>
<td></td>
<td>( m )</td>
<td>( c )</td>
<td>( m )</td>
</tr>
<tr>
<td>Pair 1</td>
<td>10.367</td>
<td>5513</td>
<td>10.268</td>
</tr>
<tr>
<td>Pair 2</td>
<td>10.378</td>
<td>5513</td>
<td>10.284</td>
</tr>
<tr>
<td>Pair 3</td>
<td>10.484</td>
<td>5519</td>
<td>10.339</td>
</tr>
<tr>
<td>Pair 4</td>
<td>10.489</td>
<td>5523</td>
<td>10.351</td>
</tr>
</tbody>
</table>

*Table 4.1: Marker calibration data, based on trend lines fitted to the graphs in Figure 4.3; \( m \) is the gradient and \( c \) is the intercept*

Applying this averaged calibration data to the individual marker sheets will typically result in an error in breast thickness measurement of 0.2mm over a range of total thicknesses of 20 - 110mm, with a maximum error in thickness measurement of 0.4mm. This is similar to the result of 0.6mm quoted by Marchant [37], based on an estimate of 5 pixels for the accuracy of locating the marker centres. Smith [36] found an error of 0.9mm but this assumed the same calibration data for each marker pair. Even with the use of the jig described above, it is thought that the best precision achievable in aligning the markers would be 0.5mm. The accuracy with which the compressed breast thickness can be measured is one of the major strengths of the Manchester Method.

Additionally, multiple pairs of magnification markers enable the measurement of breast thickness at a number of points, thereby accounting for tilt of the paddle. This can be significant, as described in Section 4.2.3. Failure to account for paddle tilt has been shown by Malkov et al [157] to be the greatest source of error in volumetric percentage breast density measurement.
4.2.3 Compression paddle tilt

Some compression paddles are specifically designed to tilt as the manufacturers believe that better contact with the breast can be achieved, with less discomfort for the woman under examination. The paddles on the units used in the feasibility study were not tilting paddles but, due to the shape of the breast, it is inevitable that there will be some tilt in the chest wall to nipple direction, as illustrated in Figure 4.5. Factors that could affect paddle tilt include applied compression force, breast composition and the model of mammography system and paddle.

![Figure 4.5: Squash ball (left) and silicon-filled bra implant (right) compressed to 100N on the GE 800T unit to demonstrate paddle tilt](image)

In order to assess the extent of the paddle tilt, the thickness of tissue in the compressed breast region (i.e. excluding the breast margin) was calculated from the digitised mammograms of 20 women who had taken part in the previously described lifestyle study [150]. The images were acquired on a Lorad M-IV mammography unit, using a rigid compression paddle. Both craniocaudal (CC) and medio-lateral oblique (MLO) views were included. A profile was drawn from the chest wall to the nipple and the breast thickness along this profile was plotted. The maximum, minimum and mean values were compared with the indicated digital reading provided by the mammography unit.

Figure 4.6 shows the variation in compressed breast thickness in five breasts covering a range of thicknesses. These plots are for the left CC view and the profiles go from the chest wall to the inner edge of the breast margin. The gradient of the plots indicates that there is paddle tilt in this direction and also that the angle of tilt is not constant for all breasts.

The mean compressed breast thickness in the sample was 55.6mm (range 33.0 – 96.3mm). The mean difference between the maximum and minimum thickness in the compressed breast region was 10.8mm (range 2.5 – 21.2mm).
Figure 4.6: Variation in compressed breast thickness from the chest wall to the inner edge of the breast margin

Figure 4.7 illustrates the relationships between this difference and the system-indicated compressed breast thickness and applied compression force.

Figure 4.7: Relationship between the difference in breast thickness over the compressed breast region and system-indicated breast thickness (left) and compression force (right)

These associations are not examined further here but warrant additional investigation as they could potentially be used to generate calibration data which relates system-indicated values to compressed breast thickness and paddle tilt, thereby removing the need for magnification markers. These values would be available in the DICOM header for digital mammography units.

The system-indicated breast thickness agreed most closely with the minimum measured breast thickness. The mean difference between minimum measured and indicated thickness was 0.8mm (range -4.0 – 8.9mm). The mean difference between the average measured thickness and
indicated thickness was 6.3mm (range -1.5 – 15.1mm). This has implications for the estimation of volume and consequently breast density. It was found that on average, using the indicated breast thickness in the volume calculation would result in an underestimation in volume of 10.5% (range -3.9 – 22.6%), compared to using the average breast thickness [143].

Although compression paddles are constructed of relatively rigid polycarbonate material with a thickness of 1.7 – 2.8mm [148], it is possible that they could bend laterally under the compression forces applied during mammography, typically 100 – 200N. The extent of lateral deformation was examined by Smith [36] who constructed a separate marker sheet with four pairs of markers at the chest wall and nipple edges. Thirty mammograms were taken with these markers in place. Following digitisation, breast thickness was measured at the four positions on the film corresponding to the marker locations. Results are shown in Figure 4.8 for the average breast thickness (solid blue line) and the minimum and maximum breast thicknesses (dotted red lines).

Breast thickness was found to be approximately constant in the lateral direction, with a standard deviation from the average thickness of ±0.7mm [36]. Given that Smith estimated the measurement uncertainty in compressed breast thickness to be ±0.9mm, deformation of the compression paddle in the lateral direction was considered insignificant in comparison to paddle tilt and is therefore not accounted for in the current implementation of the Manchester Method.

4.3 Stepwedge design

The original stepwedge [36] used in the Manchester Method was constructed of PTFE (polytetrafluoroethylene, commonly known as Teflon). This material was chosen because it has a similar mass attenuation coefficient to breast tissue, but a higher density than most plastics. This would enable a wide range of attenuation to be achieved without requiring too great a thickness, and beam hardening effects would be minimal.
The stepwedge had 25 steps, each of height 1mm and length 5mm, giving a maximum height of 25mm and a total length of 125mm. The width of all steps was 12mm. Height, width and length are defined in Figure 4.9.

![Figure 4.9: Definition of stepwedge height, width and length](image)

It was necessary to shield the sides of this wedge with lead to ensure that only those parts of each step where x-rays had travelled through the whole thickness of the wedge were imaged. Without lead shielding the image became blurred by x-rays that only passed through part of the wedge, causing the intensity in grey level to vary across each step as illustrated in Figure 4.10.

![Figure 4.10: With no shielding, the grey level across each step varies due to x-rays passing through different thicknesses of stepwedge material (left); lead shielding ensures that only those parts of the step where the x-rays have travelled through the full thickness are imaged (right)](image)

X-rays are emitted from the target (anode), located above the chest wall edge of the breast support platform. The further the stepwedge is placed from this edge (for example on the 24x30cm support platform), the more oblique the beam angle and the more significant this shadowing effect becomes, thereby reducing the usable width of each step.

When the stepwedge was used in a clinical setting, Marchant [37] found that at the higher exposure factors required for thicker and denser breasts, the 25mm stepwedge did not adequately cover the range of optical densities observed within breast tissue. An additional 10 steps were
added, increasing the height to 35mm and the length to 175mm. Because of the shadowing effect, the additional height resulted in the need to increase the width from 12mm to 15mm, giving a usable width of 4mm at height 35mm.

Despite fulfilling the theoretical requirements to enable measurement of breast density, it was found that in practice the 35mm stepwedge had too great an area to fit alongside larger breasts and was too high to be imaged alongside breasts with compressed thickness less than 35mm. The lead lining made it a relatively heavy, unwieldy device that could not easily be attached to the breast support platform. A ‘hook and eye’ method was used but this required the stepwedge to be re-positioned between carrying out the left and right MLO images, as this view requires the breast support platform to be angled at 45 - 50°. Breast density analysis requires accurate identification of two steps in the digitised image. Typically, one end of the wedge was overexposed and the other underexposed, so finding the ends and correctly identifying the steps proved difficult.

For the reasons described above, a new stepwedge was required to enable the Manchester Method to be applied in the screening programme. Sections 4.3.1 – 4.4 represent original work carried out during this PhD.

4.3.1 Stepwedge material

In order to reduce the physical dimensions of the stepwedge, it was necessary to use a material with a higher atomic number. Aluminium and magnesium were initially considered but compounds were not investigated to avoid problems with inhomogeneity. The grey level within each step is measured as the mean pixel value within a region of interest (ROI). There were concerns that the standard deviation associated with this measurement would have been too great in compounds.

Table 4.2 shows the density (ρ) and mass attenuation coefficient (μ/ρ) of potential stepwedge materials and a comparison with breast, adipose and glandular tissue [142, 166].

<table>
<thead>
<tr>
<th></th>
<th>ICRU Report 44 Definitions [142]</th>
<th>National Institute of Standards and Technology Definitions [166]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Breast</td>
<td>Adipose</td>
</tr>
<tr>
<td>ρ (g cm⁻³)</td>
<td>N/A</td>
<td>0.95</td>
</tr>
<tr>
<td>μ / ρ (cm²g⁻¹)</td>
<td>0.689</td>
<td>0.568</td>
</tr>
</tbody>
</table>

Table 4.2: Properties of the materials considered for the new stepwedge compared to PTFE and breast tissue. All attenuation coefficient values are quoted at 20keV.
Figure 4.11 shows the mass attenuation coefficients of these materials over the energy range used in mammography, typically 25 – 35keV.

![Figure 4.11: Comparison of mass attenuation coefficients of aluminium, magnesium, breast tissue and PTFE over the energy range used in mammography](image)

Magnesium has a slightly lower density and mass attenuation coefficient than aluminium but was discounted for safety reasons as it is classed as a severe irritant and is also highly flammable!

There were initial concerns that the increment in step height required for an aluminium stepwedge would be too small to be machined with any degree of accuracy. For this reason, consideration was briefly given to constructing the stepwedge from a series of aluminium foils glued together. However, this device would need to be attached to a rigid Perspex base. A stepwedge of this design would therefore be constructed of three materials: aluminium, glue and Perspex. This could result in poor reproducibility. Fortunately, it was possible to machine a stepwedge from a single block of aluminium using 0.2mm increments in step height.

### 4.3.2 Design of a prototype aluminium stepwedge

A single prototype aluminium stepwedge was constructed and evaluated [135], in order to determine the maximum, minimum and incremental step heights required, as well as the optimum step length and width. The steps taken to determine the required height range are described as follows. A 28kV molybdenum spectrum with 0.03mm molybdenum filtration was generated using specialist software [50]. This represents the most commonly selected beam quality for film-screen mammography. Although the mean photon energy of the spectrum was 16.3keV, photon attenuation by a selection of materials was considered at 20keV. Using published values of density and mass attenuation coefficient [142, 166], the corresponding thickness of aluminium which would give the same photon attenuation as varying thicknesses of
fatty and glandular tissue was calculated by rearranging equation (4.4), using aluminium as the stepwedge material.

The thickness of aluminium \( x_{Al} \) required is given by Equation (4.8):

\[
x_{Al} = \frac{(\mu_f x_f + \mu_g x_g)}{\mu_{Al}}
\]

where \( \mu_{Al} \) is the linear attenuation coefficient of aluminium (calculated using the published values of \( \frac{\mu}{\rho} \) and \( \rho_{Al} \) at 20keV).

Table 4.3 shows the theoretical thicknesses of aluminium required to produce the same attenuation as a range of breast thicknesses and compositions. The results suggest that the aluminium stepwedge requires minimum and maximum step heights of 1mm and 8mm respectively, with step increments of 0.2mm.

<table>
<thead>
<tr>
<th>Material thickness (mm)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>Adipose</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>60</td>
<td>20</td>
</tr>
<tr>
<td>70</td>
<td>20</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 4.3: Theoretical thickness of aluminium required to produce the same x-ray photon attenuation as a range of breast thicknesses at mean photon energy 20keV

The design of the prototype aluminium stepwedge is shown in Figure 4.12. Lead lining was not considered necessary as the smaller dimensions of the wedge, particularly height, would reduce the problem of shadowing. The usable width of each step should be greater for the same reason. A width of 10mm was chosen. The step length was varied from 3mm to 5mm and the increment in step height was also varied between 0.2mm and 1mm. This was primarily to ease with the construction of the stepwedge. At this stage of development, it was only considered necessary to assess the adequacy of the incremental height over a small range of step heights.
Figure 4.12: (a) side view and (b) bird’s eye view of aluminium prototype showing the variation in step height and width and the location of notches to help identify each step.

An important practical development was the introduction of notches to aid with step identification on the digitised x-ray image. Figure 4.13a shows a comparison of the aluminium prototype with the lead-lined PTFE stepwedges, and Figure 4.13b shows how the 35mm PTFE stepwedge and the aluminium prototype would fit alongside a 40mm Perspex breast phantom.

Figure 4.13: (a) Comparison of stepwedges. Top to bottom: 35mm PTFE, 25mm PTFE and aluminium prototype; (b) Stepwedges of 35mm PTFE (top) and aluminium (bottom) alongside 40mm Perspex.
4.3.3 Evaluation of the prototype aluminium stepwedge

A number of investigations were carried out on a Lorad M-IV mammography unit to determine if the stepwedge dimensions were adequate. Measurements were made using the 18x24cm and 24x30cm breast support platforms. The prototype aluminium stepwedge was placed on the breast support platform at the location indicated in Figure 4.13b and was imaged alongside 2, 4, and 6cm of Perspex, compressed to 50N. Because glandularity typically decreases with breast thickness, these Perspex thicknesses are thought to represent a range of breast thicknesses from 2.1 to 7.5cm [167]. The clinically used setting (Auto Filter) was employed so that exposure factors would be representative of those selected in clinical practice. ‘Auto Filter’ is an advanced automatic exposure control (AEC) mode which selects the appropriate kV based on the indicated breast thickness and the appropriate target / filter and mAs based on a very short pre-exposure. The dose-rate reaching the AEC detector during the pre-exposure gives an indication of the density of the breast. All exposures were repeated using the 35mm PTFE stepwedge placed in the same location, as shown in Figure 4.13b.

The films were digitised using a Kodak LS85 digitiser at a pixel size of 50 μm and a pixel depth of 12 bits (4096 grey levels). The grey level is linearly related to optical density (OD) in the range 0.03 - 4.10 OD [168]. For this particular digitiser, grey level 0 corresponds to an OD of zero (white) and grey level 4096 corresponds to the maximum OD value (black). The pixel depth was reduced to 8 bits (256 grey levels) after digitising, using a window based on the maximum and minimum OD present in the image. This was to reduce file sizes of the stored images. The images were analysed using ImageJ, a free Java-based software application.

Step height was assessed by comparing the PTFE and aluminium stepwedges imaged alongside 4cm Perspex under identical exposure conditions (25kV, Mo/Mo, 135mAs). The optical density of each step on the aluminium and PTFE stepwedge images was measured using a Parry Transmission Densitometer DT1105 with a 1.0mm aperture. Results are shown in Figure 4.14.

For step increments of 0.2mm in the aluminium stepwedge, the change in optical density (OD) per step was equal to or less than that in the PTFE stepwedge, suggesting that a step increment of 0.2mm is satisfactory. The step increment of 1mm is certainly not adequate from step heights of 3 to 5mm but may in fact be suitable at greater step heights.
Figure 4.14: Optical density versus step height for aluminium and PTFE stepwedges

Under the exposure conditions for 4cm Perspex, the image of the prototype aluminium stepwedge was acceptable with optical densities from 0.25 to 3.70 being covered (Figure 4.15).

Figure 4.15: Aluminium prototype stepwedge exposed using 25kV, Mo/Mo, 135mAs

However, Figure 4.16 shows that at the higher exposure factors selected for 6cm Perspex (29kV, Mo/Rh, 241mAs), the lowest optical density in the stepwedge was 1.0. Conversely, for the lower exposure factors required for 2cm Perspex (24kV, Mo/Mo, 32mAs), the highest optical density was 2.4.

The extent of this problem is analysed with reference to Figure 4.17. Although the characteristic curve for film typically covers an optical density range of 0.25 OD (base plus fog) to 4.0 OD ($D_{\text{max}}$), it is unlikely that the image of the breast will include such an extensive range.
The blue line in Figure 4.17 shows the characteristic curve for Kodak Min R-2000 film [169]. The red dotted lines indicate the optical densities that are expected in a correctly exposed mammogram. All diagnostically significant details should be imaged with a density between 0.6 and 2.2, except for the subcutis and microcalcifications [170].
Based on these results, the height dimensions of the prototype stepwedge are considered inadequate to sufficiently cover the range of optical densities within breast tissue over the full range of exposures expected clinically.

Note that the unexpected upturn in the curves in at the maximum stepwedge thickness of 8 mm in Figure 4.16 has been observed in previous work using the PTFE stepwedge [37]. It is thought to be an artefact, attributed to the fact that on the image, only a very small area of the last step is visible for which the x-rays have passed through the whole stepwedge thickness, with the remainder of the step appearing as heavily shadowed.

Step width was assessed by analysing line profiles of pixel value across the width of each step within the stepwedge exposed alongside 4cm Perspex at 25kV, Mo/Mo, 135mAs (Figure 4.18). Step width was 200 pixels. With increasing step height, it became progressively more difficult to differentiate the step edges from the background so the line profiles covered a greater distance than just the step width and sharp changes in pixel value were used to determine the edges of the step.

Because the stepwedge is not lead-lined, there is a shadowing effect at the edges which becomes more significant as step height increases and the shape of these profiles reflects those in Figure 4.10 (left). However, within each step, there is a relatively uniform region where fluctuation in pixel value is low. The step width is 10mm and Figure 4.18 suggests that, at best, only the central 5mm (pixel distance 50 – 150) can be used to determine the true mean pixel value within the step.
The optimum size of the region of interest (ROI) used in the analysis software was investigated as follows. Mean pixel value (signal) and standard deviation (noise) were measured in ROIs of varying area and position drawn within all steps on the stepwedge exposed alongside 4cm Perspex at 25kV, Mo/Mo, 135mAs. The ROIs are shown in Figure 4.19; (a) covered 1000 pixels (50 x 20), (b) covered 2000 pixels (100 x 20) and (c) covered 4000 pixels (100 x 40). The effects of ROI size, step height and step length on mean pixel value (MPV) and standard deviation (Std Dev) are shown in Table 4.4 for regions (a), (b) and (c).

![Figure 4.19: ROI size and position](image)

<table>
<thead>
<tr>
<th>Step height (mm)</th>
<th>Step width (mm)</th>
<th>ROI (a) MPV</th>
<th>Std Dev</th>
<th>ROI (b) MPV</th>
<th>Std Dev</th>
<th>ROI (c) MPV</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>3</td>
<td>227.76</td>
<td>2.19</td>
<td>227.47</td>
<td>2.28</td>
<td>227.40</td>
<td>2.37</td>
</tr>
<tr>
<td>1.2</td>
<td>4</td>
<td>223.08</td>
<td>2.42</td>
<td>223.07</td>
<td>2.51</td>
<td>222.98</td>
<td>2.49</td>
</tr>
<tr>
<td>1.4</td>
<td>5</td>
<td>216.62</td>
<td>2.44</td>
<td>216.37</td>
<td>2.51</td>
<td>216.37</td>
<td>2.56</td>
</tr>
<tr>
<td>1.6</td>
<td>3</td>
<td>207.40</td>
<td>2.05</td>
<td>207.39</td>
<td>2.15</td>
<td>207.46</td>
<td>2.18</td>
</tr>
<tr>
<td>1.8</td>
<td>4</td>
<td>194.66</td>
<td>1.97</td>
<td>194.95</td>
<td>2.05</td>
<td>194.89</td>
<td>1.94</td>
</tr>
<tr>
<td>2.0</td>
<td>5</td>
<td>179.69</td>
<td>1.68</td>
<td>179.87</td>
<td>1.84</td>
<td>179.78</td>
<td>1.92</td>
</tr>
<tr>
<td>2.2</td>
<td>3</td>
<td>158.55</td>
<td>2.27</td>
<td>158.99</td>
<td>2.45</td>
<td>159.06</td>
<td>2.54</td>
</tr>
<tr>
<td>2.4</td>
<td>4</td>
<td>136.75</td>
<td>2.03</td>
<td>137.56</td>
<td>2.23</td>
<td>137.49</td>
<td>2.41</td>
</tr>
<tr>
<td>2.6</td>
<td>5</td>
<td>110.95</td>
<td>2.24</td>
<td>112.31</td>
<td>2.83</td>
<td>112.5</td>
<td>2.81</td>
</tr>
<tr>
<td>2.8</td>
<td>3</td>
<td>86.43</td>
<td>2.13</td>
<td>88.14</td>
<td>2.98</td>
<td>88.12</td>
<td>2.96</td>
</tr>
<tr>
<td>3.0</td>
<td>4</td>
<td>68.21</td>
<td>1.81</td>
<td>69.48</td>
<td>2.52</td>
<td>69.40</td>
<td>2.58</td>
</tr>
<tr>
<td>4.0</td>
<td>5</td>
<td>26.50</td>
<td>0.63</td>
<td>27.38</td>
<td>1.34</td>
<td>27.41</td>
<td>1.36</td>
</tr>
<tr>
<td>5.0</td>
<td>3</td>
<td>20.12</td>
<td>0.32</td>
<td>20.61</td>
<td>0.77</td>
<td>20.56</td>
<td>0.64</td>
</tr>
<tr>
<td>5.2</td>
<td>4</td>
<td>19.91</td>
<td>0.29</td>
<td>20.14</td>
<td>0.48</td>
<td>20.16</td>
<td>0.48</td>
</tr>
<tr>
<td>5.4</td>
<td>5</td>
<td>19.45</td>
<td>0.50</td>
<td>19.74</td>
<td>0.56</td>
<td>19.75</td>
<td>0.56</td>
</tr>
<tr>
<td>5.6</td>
<td>3</td>
<td>19.06</td>
<td>0.24</td>
<td>19.44</td>
<td>0.53</td>
<td>19.46</td>
<td>0.52</td>
</tr>
<tr>
<td>5.8</td>
<td>4</td>
<td>19.02</td>
<td>0.23</td>
<td>19.29</td>
<td>0.45</td>
<td>19.29</td>
<td>0.47</td>
</tr>
<tr>
<td>6.0</td>
<td>5</td>
<td>19.01</td>
<td>0.27</td>
<td>19.21</td>
<td>0.45</td>
<td>19.20</td>
<td>0.41</td>
</tr>
<tr>
<td>7.0</td>
<td>3</td>
<td>19.00</td>
<td>0.26</td>
<td>19.09</td>
<td>0.29</td>
<td>19.12</td>
<td>0.32</td>
</tr>
<tr>
<td>8.0</td>
<td>4</td>
<td>20.97</td>
<td>0.79</td>
<td>21.81</td>
<td>1.03</td>
<td>21.78</td>
<td>1.81</td>
</tr>
</tbody>
</table>

Table 4.4: Effect of step height and ROI area on signal (MPV) and noise (Std Dev)
A number of observations can be made. Regardless of ROI size, the standard deviation remains approximately constant to a step height of 3.0mm. Fluctuations are random with step height (reflected in the profiles in Figure 4.18) and also with step length, suggesting that the optimum step length is 3mm in order to make the step wedge as compact as possible. Standard deviation drops sharply from 3.0 to 5.2mm and then remains approximately constant. The reason for this is likely to be that these thicknesses of aluminium are so attenuating under these exposure conditions, that there are almost no x-ray photons reaching the film. The mean pixel value is very low and there is little variation in this low signal. For comparison, ROIs of size 1000 pixels and 10000 pixels were drawn in the film identification label, which represents an area of no signal (or base plus fog). The MPV and standard deviation were (18.343, 0.475) and (18.524, 0.501) for the 1000 and 10000 pixel areas respectively. These agree well with the results in Table 4.4 for step heights of 5.0mm and above using the smallest ROIs, (a) and (b).

The size of the ROI has no effect on MPV but standard deviation increases slightly with ROI size. Note that care was taken to position these ROIs to avoid the edge effect demonstrated in Figure 4.18. An ROI size of 50 x 20 pixels equates to a distance of 2.5 x 1.0mm. A different digitiser was used in the feasibility study which had a pixel resolution of 44\( \mu \)m and the ROI size used in the analysis software was 2.20 x 1.32mm.

### 4.3.4 Stepwedge suitable for clinical use

Because the range of step heights was inadequate, a second prototype was designed with heights from 0.4mm to 12mm. Step increment was 0.2mm from 0.4 to 6mm and 0.5mm from 6 to 12mm. Consideration was given to the number of steps required to achieve a sufficient number of points to create a smooth curve but placing an upper limit on the overall length of the step wedge, hence the different step increments. This step wedge was imaged alongside 2, 4 and 6cm Perspex on the same Lorad M-IV mammography unit as used previously, again using Auto Filter mode. For the exposure factors used for 2cm and 6cm Perspex, the maximum and minimum optical densities were 3.85 and 0.54 respectively. However, given that compressed breast thickness can exceed 10cm, it was decided that a greater maximum height would result in fewer failures of the method. The minimum height of 0.4mm was satisfactory. The curves were smooth, suggesting that a coarser step increment above 6mm is acceptable.

A third prototype was designed with step heights from 0.4mm to 14mm and step lengths of 3mm. Step increment was 0.2mm from 0.4 – 6mm, 0.5mm from 6 – 8mm and 1mm from 8 – 14mm. Notches were introduced every mm. Figure 4.20 shows a birds-eye view and a side view of this prototype step wedge. It was imaged alongside 2, 4 and 6cm Perspex on the same Lorad M-IV mammography unit, using Auto Filter mode. Figure 4.21 shows the resulting plots of optical density versus step height.
Figure 4.20: Third aluminium prototype step wedge: bird’s eye view (top) and side view (bottom)

Figure 4.21: Optical densities on the third prototype aluminium step wedge exposed alongside 2, 4 and 6cm Perspex. Exposure factors are shown in the legend.

The range of optical densities covered means that this prototype is thought to be acceptable for clinical use. For the exposure factors required for 6cm Perspex, the minimum optical density is 0.50. It is interesting to note that the upturn artefact is no longer present. An extra 10mm has been added to the final step, giving a total length of 127mm, in order to facilitate a mechanism for attaching the step wedge to the breast support platform. The GE mammography units used in the
feasibility study had a neat mechanism for attaching radio-opaque view markers to the breast support platform. This was adapted for the stepwedge as shown in Figure 4.22a.

Figure 4.22: (a) Stepwedge attached to breast support platform; (b) stepwedge remains securely in place during MLO projections; (c) stepwedge can be easily flipped out of the field of view

The stepwedge remains securely in place during both the craniocaudal (CC) and mediolateral oblique (MLO) views, which require the x-ray tube and breast support platform to rotate to an angle of 45 - 50°, Figure 4.22b. This has advantages over the ‘hook and eye’ method used for the PTFE stepwedge, which was essentially two paperclips. Previously, it was necessary to change the location of the stepwedge for the left and right MLO views, due to the slope of the breast support platform. This is no longer necessary. Two stepwedges were provided for each mammography unit; one to attach to the 18x24cm breast support platform and one to attach to the 24x30cm breast support platform. For both field sizes, there would be rare occasions where the stepwedge would not fit alongside the breast, due to a very large breast area. In these instances, the stepwedge could simply be flipped out of the way, as shown in Figure 4.22c. Unfortunately breast density analysis could not subsequently be performed on these images.
4.4 Stepwedge calibration

Breast tissue transmission is calibrated against phantoms of varying thicknesses of fatty and glandular tissue equivalent material using the stepwedge as a practical intermediate step. The materials used are epoxy resin based tissue substitutes AP6 and WT1 which simulate fatty and glandular tissue respectively [142]. The tissue equivalent phantoms would be too large to include alongside the breast, hence the need for the stepwedge.

The adipose-equivalent material, AP6, was developed by White et al [171] in 1977 and is composed of epoxy resin with fillers of PTFE and polyethylene. White et al [171] also developed a breast tissue substitute, BR12, which was composed of the same materials as AP6, with the addition of calcium carbonate. The composition of BR12 reflected the assumption in ICRP Publication 23 (Report of the Task Group on Reference Man, 1975) [172] that the average breast was 50% fat and 50% water. The material WT1 [173] has the same components as BR12 but in different proportions.

Hammerstein et al [174] subsequently determined the elemental composition of fat and glandular tissue within the breast and their results were used to inform the construction of the CIRS tissue equivalent material [146] used by Shepherd et al [157]. Johns and Yaffe [47] and Byng et al [175] have also carried out experimental work to determine the density and linear attenuation co-efficient of breast tissue and tissue equivalent materials using a technique known as pulse-height spectroscopic analysis.

Comparisons of density and linear attenuation coefficient are shown in Table 4.5 for breast tissue and Table 4.6 for tissue equivalent materials. Linear attenuation coefficient is energy dependent so is quoted for 20keV.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ρ (gcm⁻³)</td>
<td>μ (cm⁻¹)</td>
<td>ρ (gcm⁻³)</td>
</tr>
<tr>
<td>Adipose</td>
<td>0.950</td>
<td>0.471</td>
<td>0.930</td>
</tr>
<tr>
<td></td>
<td>(0.917 – 0.939)</td>
<td>(0.441 – 0.476)</td>
<td></td>
</tr>
<tr>
<td>Gland</td>
<td>1.020</td>
<td>N/A</td>
<td>1.040</td>
</tr>
</tbody>
</table>

Table 4.5: Comparison of density and linear attenuation coefficient for breast tissue
Table 4.6: Comparison of density and linear attenuation coefficient for tissue equivalent material

Although WT1 is commonly known as “solid water”, it can be seen that the density exactly matches that of glandular breast tissue quoted in ICRP Report 23 [172]. The linear attenuation coefficient of 0.838 cm\(^{-1}\) is somewhat higher than that of CIRS gland (0.747 cm\(^{-1}\)), which reflects that of Hammerstein et al (0.735 cm\(^{-1}\)) [174]. However, it is actually a better reflection of the more recent data derived experimentally by Johns and Yaffe [47], so is therefore considered an acceptable material for simulating glandular tissue.

Calibration data were collected for the two units used in the feasibility study: a GE Senographe DMR at Bury and a GE Senographe 800T on Bolton Mobile Van 2. Stepwedge calibration was carried out under automatic exposure control (Standard Auto mode) as this is how the unit is used clinically. In this mode, the unit automatically selects the parameters to achieve the optimum balance between image quality and dose, based on in-built look-up tables of breast thickness and a very short pre-exposure.

The stepwedge was imaged alongside phantoms with total compressed thickness ranging from 2 cm to 7 cm to represent the typical range encountered clinically. At least three combinations of fatty and glandular tissue equivalent material were used to make up each total breast thickness, always including 100% fatty. Films were digitised using a Vidar CadPro Advantage digitiser using a pixel size of 44 \(\mu\)m and a pixel depth of 16 bits (65536 grey levels), as the only options available were 8 or 16 bits. However, they were converted to 12-bit images (4096 grey levels) prior to analysis. This digitiser covers a range of optical densities from 0.05 to 4.20 and uses a linear translation table [171]. Following digitisation, software written in Matlab (The Mathworks, Inc) was used to calculate the thickness of aluminium with the same grey level as a region of interest (ROI) drawn within the tissue equivalent phantom. The ROI was 10x10 mm and was positioned 4 cm back from the chest wall as this is where optical density measurements are made during mammography quality control tests [164]. Graphs of stepwedge thickness versus glandular tissue thickness were plotted for each total phantom thickness. Results are shown in Figure 4.23 for the 18x24 cm field size at Bury and Bolton.

Linear interpolation between data points and extrapolation to greater breast thicknesses was carried out in Matlab to generate a calibration data surface shown in Figure 4.24, which conveyed the relationship between stepwedge thickness, glandular tissue thickness and breast...
thickness. Breast thickness and equivalent stepwedge thickness are known for every column of tissue projected onto a pixel in the mammogram. This calibration surface is therefore used to calculate glandular tissue thickness at each pixel.

Figure 4.23: Stepwedge thickness versus glandular thickness; legend shows total breast thickness
Data collected on (a) GE 800T at Bolton; (b) GE DMR at Bury

Data are closely matched for breast thicknesses of 20 – 50mm and for low glandularities at breast thicknesses of 60 and 70mm. However, it can be seen that a greater stepwedge thickness is required for thick, dense breasts imaged at Bolton, compared to Bury. One reason may be that the GE 800T unit at Bolton only has a molybdenum target, with molybdenum and rhodium filters. The GE DMR unit at Bury has an additional rhodium target which provides a harder, more penetrating x-ray beam and therefore requires reduced mAs under automatic exposure control. It is also possible that, because the stepwedge is not tissue equivalent, there is a difference in the energy dependence of the attenuation coefficients, which is causing a change in matching thickness at different energies.
Figure 4.24: Calibration surface generated in Matlab, The surface relates stepwedge thickness, glandular thickness and total breast thickness

There is a possibility that the stepwedge may not be adequate when exposed alongside the thickest, densest breasts. It is anticipated that there will only be a small proportion of incidences when the method fails for this reason as breast glandularity typically decreases with increasing breast thickness, as shown in Table 4.5 [167]. Note that these values are for the central portion of the breast as surface layers of adipose tissue 5mm thick are assumed. This would imply that the overall percentage volumetric breast density is in fact even lower.

<table>
<thead>
<tr>
<th>Compressed breast thickness (cm)</th>
<th>Glandularity age 50 – 64 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 4.7: Typical glandularity variation with breast thickness for the screening population [167]
Figure 4.25 shows the optical density profiles of the stepwedge imaged alongside a range of breast thicknesses and compositions.

![Figure 4.25: Optical density versus stepwedge thickness](image)

The legend shows the total breast thickness with percentage glandularity in brackets.

It is not possible to recreate the glandularities from Table 4.5 exactly, as the tissue equivalent material comes in 0.5cm or 1cm slabs. Assuming a glandularity of 17% for a 6cm phantom (1cm glandular) and 14% for a 7cm phantom (1cm glandular), it can be seen that the stepwedge reaches minimum optical densities of 0.45 and 0.80OD respectively and the method is likely to be successful. However, for large breasts that are highly dense (greater than 50%), the method is likely to fail. The feasibility study (Chapter 7) will indicate the extent of this problem.

### 4.5 Breast margin

The breast is considered to have a central compressed breast region and a margin. Breast thickness measurement in the compressed region has been described in detail in Section 4.2. Breast thickness measurement in the margin is far more complex. The outer edge of the breast margin must first be detected in order to segment the breast from the background. The inner edge of the margin must then be located; this represents the point at which the breast loses contact with the compression paddle. Finally, the thickness within the margin must be estimated. This is not constant, but decreases rapidly in the direction from inner to outer edge.
4.5.1 Previous approach

Marchant [37] developed a method for locating the inner and outer edges of the breast margin, based on work by Byng et al [177] who used the technique in developing a thickness equalisation algorithm. The reduced breast thickness in the margin results in greater x-ray transmission and therefore reduced brightness, making it more difficult to assess subtle contrasts. Increasing the brightness in the breast margin to match that in the central region would improve image interpretation.

The first step towards locating the breast margin is to smooth (blur) the image to suppress variations in grey level associated with composition. Marchant [37] used a Gaussian filter with a radius of 2.5mm to achieve this. Within the analysis software, the user is first asked to define an approximate breast edge. This is shown in blue on Figure 4.26. Starting at this location, horizontal profiles of grey level are plotted for every row, as indicated by the yellow line. The resulting profile is shown in Figure 4.27. If the image has been sufficiently smoothed, changes in grey level should be the result of changes in breast thickness only. The outer edge of the breast margin should be easily determined as the point at which the grey level first decreases significantly below the background value.

![Image of breast image smoothed using a Gaussian filter](image)

*Figure 4.26: Breast image smoothed using a Gaussian filter*

*Blue: User-defined approximate breast outline; Yellow: horizontal profile, plotted in Figure 4.27*

The inner edge is deemed by Marchant [37] to be the point at which the grey level stops falling. However, there are limitations to this approach as the Gaussian filter applied to the image does not completely eliminate variations due to breast composition. It would be very difficult to achieve a completely uniform grey level in a breast with large regions of dense tissue, such as that in Figure 4.26. Rather than looking for the point at which the grey level stops falling, it may be more appropriate to examine changes in the grey level gradient.
Marchant [37] extended this method to estimate the breast thickness in the margin. The average grey levels at the inner and outer edges were determined and all profiles were scaled to obtain grey levels between these values. Additionally, their length was adjusted to match the average breast margin width. The average grey level profile was converted to a tissue thickness profile with reference to the stepwedge image. Unfortunately, this method resulted in significant thicknesses of dense tissue measured close to the breast edge where the breast is known to be predominantly fat. An additional limitation was that assuming a constant margin width proved problematic in locations where the breast edge was parallel to the horizontal profiles. It was thought that radial profiles (Figure 4.28), would overcome some of these limitations.

Patel [163] investigated this approach using profiles spaced at 5°. However, locating the central point from which the profiles would radiate was non-trivial. This point is shown by the pale blue circle in Figure 4.28. An automated approach was trialled which detected the two breast edge points based on a grey level profile along the chest wall edge. The central location was taken to be the midpoint. However, fluctuations in grey level meant that the breast edge did not occur at a distinct pixel, but over a range of pixels, making it impossible to accurately determine the
central point. User input was therefore required to select the breast edge points. However, the accuracy of this measurement is not guaranteed as it can be very difficult to mark the exact pixel at which this occurs. Although there was some success in using this method on CC views, it could not be applied to MLO views due to the inclusion of pectoral muscle.

When analysing images from the lifestyle study [150], Marchant [37] and Patel [150] assumed a constant breast thickness in the margin, equal to the compressed breast thickness at the inner edge of the margin. They acknowledged that this would overestimate breast thickness and therefore underestimate breast density within the margin. However, given that the breast margin contains mostly fatty tissue, this approach was considered preferable to overestimating breast density within this region.

### 4.5.2 Current approach

The use of horizontal grey level profiles to detect the outer breast edge, developed by Marchant [37], was initially employed in the analysis of images from the feasibility study. However, defining an approximate breast edge proved to be time consuming and was found to be a major source of inter-observer variation by Verow et al [178]. Selecting a threshold to separate the breast from the background is highly dependent upon the noise present in the image. A preferable solution would be to use this as the initial input to an active contour model algorithm. Such a technique was developed by Michael Berks (research associate at ISBE) with the intention of applying it to a system for computer aided diagnosis (CAD) of breast cancer. It required only minor modification to make it suitable for detecting the breast edge in the density analysis software and had the huge advantage of being fully automated. The method is based on that by Ferrari et al [179] and is described as follows, with reference to Figure 4.29.

*Figure 4.29: (a) original mammogram; (b) & (c) binary image before and after applying morphological operators*
Firstly, dynamic range compression is applied which enhances the contrast of the regions near the breast boundary where details are poorly defined. A global threshold based on analysis of the grey-level histogram in each mammogram is then applied which segments the breast and other markers from the background, Figure 4.29b. Morphological operators are applied to the resulting binary image to isolate the main breast region and smooth the edge, Figure 4.29c. The approximate breast edge contour is used as the input to an adapted active contour algorithm which then computes a more precise and locally smooth demarcation of the breast edge.

The Manchester Method currently incorporates an elliptical model of breast thickness estimation within the breast margin, described with reference to Figure 4.30.

![Figure 4.30: Side view of the breast in the CC position](image)

At the outer edge of the breast margin (i.e. nipple location) the vertical distance between the compression paddle and the breast support platform is \( d_1 \). This vertical distance is known from the breast thickness profile generated by the magnification markers. Radius \( r_1 \) is then defined as \( (d_1) / 2 \). The inner edge of the breast margin is assumed to lie at a distance \( r_1 \) from the nipple, moving in the chest wall direction. This corresponds to the point at which the breast loses contact with the compression paddle. The vertical distance between the compression paddle and the breast support platform is \( d_2 \), with radius \( r_2 = (d_2) / 2 \). The thickness in the breast margin is modelled as an ellipse with horizontal radius \( r_1 \) and vertical radius \( r_2 \).

This process is repeated at a number of points around the breast edge as seen on the mammogram (red dots in Figure 4.31a), with \( r_1 \) and \( r_2 \) computed for each point. The inner edge of the breast margin is determined in the direction normal to the outer breast edge at that point (green lines in Figure 4.31a). Smoothing between the elliptical profiles computed at each point then takes place to give the model of the breast edge (Figure 4.31b).
Figure 4.31: (a) red dots indicate points on the breast edge and green lines show the normal profiles plotted to determine the breast margin; (b) smoothed breast edge; (c) results of thickness profiles drawn from the chest wall to the breast edge for four views from one examination.

The image in Figure 4.31b is displayed as a thickness map where white represents the maximum thickness and black is zero thickness (i.e. background). Profiles drawn from the chest wall to the nipple would have the appearance shown in Figure 4.31c. These show 4 views from one examination. The linear portion represents the thickness variation in the compressed breast region and the curved portion shows how the breast thickness decreases rapidly in the breast periphery. The curve appears to be smooth with no discontinuity. The same characteristics were observed for profiles drawn from the chest wall to the breast edge over a range of angles [143].

4.6 Analysis software

The Manchester Method is a semi-automated technique. This section describes the methodology from the point of view of an operator. The analysis software was written using Matlab 7.1 (The Mathworks, Inc.) with assistance from Michael Berks.

All films from the feasibility study were digitised using a Vidar CadPro Advantage digitiser, selecting pixel size of 44 μm and pixel depth of 16 bits (65536 grey levels). Prior to analysis, software was run on all images in order to reduce them to 12 bits (4096 grey levels) and to anonymise them. This was simply done by defining the region of the identification label and making all pixels within this region equal to 4095 (black).

Note that the default appearance of the image in Matlab is a negative of the mammogram but the grey level values remain unchanged so that 4095 represents high optical density (black) and 0 represents low optical density (white) in the original mammogram.
After selecting the mammogram images they wish to analyse, the operator is presented with an image showing how the breast has been segmented from the background. They are asked to confirm that the breast border (green line in Figure 4.32) is acceptable.

![Figure 4.32: Left: Original mammogram; Right: breast border defined by automatic segmentation](image)

For the MLO image, the user is required to manually segment the pectoral muscle (Figure 4.33).

![Figure 4.33: Left: Original mammogram; Right: user-defined pectoral muscle](image)

The operator is then asked to click on each of the markers if both markers within a pair are visible. For 18x24cm images they are asked to select the 3 pairs closest to the chest wall as the 4th pair is generally in the ID marker. For 24x30cm images, they are asked to select all 4 pairs, if present.
A “Canny” edge enhancement process [180] is used to find the edges of the marker. This takes the grey level intensity image, \( I \), as the input and returns a binary image of the same size, with ones where the function finds edges in \( I \) and zeros elsewhere. The Canny method finds edges by looking for local maxima of the grey level gradient of \( I \). The edges of the marker are not always well-defined, especially if they occur very close to the edge of the film. Two thresholds are therefore employed, to detect strong and weak edges. Weak edges are only included in the output if they are connected to strong edges, which makes this method less susceptible to noise than other edge detection algorithms.

The “Hough transform” is used to fit a circle to the marker. This transforms complex patterns of pixels in the image domain into compact features in a chosen parameter space. The technique is robust to some loss of data due to, for example, noise in the image [181]. Figure 4.35 illustrates the use of the Hough transform for circle detection.

The image domain and parameter space are congruent in this case. To detect a circle in the image domain (blue), circles of the same radius are plotted in parameter space (red), centred on the edge segments (found by the Canny edge detector). The peak in the parameter space indicates the centre of the detected circle. For this reason, the Hough transform is often referred to as a “voting system”.

Figure 4.34: Left: the user clicks on each of the markers; Right: a circle is fitted showing the centre of the marker

Figure 4.35: Hough transform for circle detection. The centre of a circle in the image domain is located by the peak signal in the parameter space [181]
If the edges of the marker are not well defined, the Canny edge detection method is unsuccessful and the user is prompted to manually select at least 3 points on the edge of the marker. A circle is then fitted based on these points, again using the Hough transform. This process is shown in Figure 4.36. In these examples, the marker is very close to the breast edge (left) and obscured by a view marker (right). It does not matter if the whole marker is not visible as long as the centre-point lies on the image.

![Figure 4.36: Examples where the Canny edge detector is unsuccessful and the user must manually select points on the edge of the marker](image)

The distance between the centres of each marker within a pair is calculated and the previously derived calibration data (Section 4.2.2) is used to convert this to breast thickness. Figure 4.37 shows the thickness profile plot for the original mammogram in Figure 4.32. Note that the operator would not normally see this plot.
Once the breast thickness has been determined, the image is resized so that the pixel size is $220 \mu m$. This is considered sufficient for the remaining analysis steps and speeds up computation.

The operator is then asked to select two steps on the stepwedge and indicate what number these steps are; the notches assist with this. Based on this input, the positions of all 39 steps are shown (Figure 4.38). Note that the actual regions selected (2.20 x 1.32mm) are smaller than these so this is schematic only. If a mistake is made, the user has the option of nudging 5 steps to the left or right since the most likely error is mis-counting the notches. Alternatively, they may re-select two steps.

Figure 4.37: Compressed breast thickness profile. Red dots: actual data points based on the magnification markers; Green line: fit to data, extrapolated across image

Figure 4.38: The user marks the edge of two steps and enters the numbers of these steps. ROIs in each step are shown so the user can check that they have correctly defined the steps
The mean pixel value within each step is determined. The mean grey level is expected to decrease with increasing step height as shown in Figure 4.39. If there has been an error, for example the stepwedge is partially obscured by the ID label, or a view marker encroaches on the stepwedge, the grey level may increase unexpectedly. If this happens, the value of the previous step is assigned instead. This plot would usually remain hidden from the operator, in order to reduce the time taken to run the software.

All remaining steps are fully automated. The inner edge of the breast margin is located and the thickness in the margin is modelled as an ellipse.

Finally, the user is presented with the glandular thickness map shown alongside the original mammogram, Figure 4.40. This is used as a visual confirmation that the method has been successful. The greyscale bar on the right shows the glandular thickness. This is a very dense breast, having percentage volumetric density of 42.0%.

An experienced operator ran the software on 187 images in 170 minutes, giving an analysis time of 55s per image. Any images which did not require user input were suppressed.
4.7 Discussion

A number of modifications have been made to the Manchester Method since its inception. The most substantial development is a new, compact stepwedge which enables the method to be applied in the breast screening environment. The marker sheets have also been re-designed. One of the main advantages of the Manchester Method is that breast thickness measurement is highly precise and is not susceptible to variations in compression force. The average error in breast thickness measurement is estimated as 0.2mm. Tolerances are expected to be tight as calibration data were generated using phantoms with thicknesses known to within 0.1mm. Additionally, the edges and centres of the markers are determined using automated techniques on images with a pixel size of only 44μm. The fact that the calibration data are averaged over two mammography units is thought to result in a maximum error of 0.4mm, which still improves upon the error of 0.9mm quoted for the original method [36]. Furthermore, using several pairs of magnification markers enables the measurement of breast thickness at a series of points thereby accounting for paddle tilt. Failure to account for tilt can lead to significant errors in density measurement [157].

The active contour model algorithm developed by Michael Berks provides more accurate detection of the breast edge than methods based solely on grey level profiles, regardless of whether these are horizontal [37] or radial [163]. This technique also has the great advantage of being fully automated.

A system for modelling tissue thickness within the breast margin has now been incorporated into the Manchester Method and was applied to clinical images for the first time during the feasibility study. Using this model is considered preferable to assuming a constant tissue thickness within
the breast margin but there are still occasions where the glandular thickness maps reveal an overestimation of the thickness of dense tissue in the margin. This represents an area where additional work is required and further discussions are presented in Chapter 10.

The main limitation with the Manchester Method is that it cannot be applied retrospectively, because of the requirement to include the stepwedge and markers in every image. For digital mammography, it may be possible to use the integrated detector as the calibration device, thereby removing the need for the stepwedge. Preliminary investigations are presented in Chapter 9. However, it would be very difficult to achieve highly accurate measurements of compressed breast thickness, which also account for paddle tilt, without using magnification markers. The current markers have some disadvantages; for example, they can be obscured by view markers or film labels, although this would not be an issue for digital mammography. The fact that they may occasionally overly the breast tissue is more concerning although it is anticipated that this would occur predominantly in the pectoral muscle or skin edge. Previous work has shown that the markers only overly tissue in 2.5% of cases and do not overly the same area of tissue on both views [182]. A possible solution may be to place the marker pairs on one corner of the compression paddle. Assuming the paddle tilts and does not bend, only two pairs of markers would be needed.

Finally, it is recognised that aluminium has a higher effective atomic number than breast tissue (Z = 13 compared with Z_{eff} \sim 7) and beam hardening effects imply that equal pixel values under the stepwedge and breast will not represent equal photon or energy fluences. This may explain why the maximum step height of 8mm was inadequate in the first prototype stepwedge. Figure 4.41 shows how a typical mammography spectrum is modified when transmitted through 5mm aluminium and 50mm soft tissue [50].

Figure 4.41: Typical mammography spectrum of 28kV, Mo/Mo transmitted through 5mm aluminium (blue line) and 50mm soft tissue (pink line) [50]
However, the stepwedge is calibrated against tissue equivalent material and merely provides a practical intermediate step in the process of grey level matching. It is not considered necessary for the stepwedge to provide the same absolute transmission as the breast or phantom, providing the relative values are the same. It is acknowledged that this will only hold true as long as the calibration x-ray spectrum and clinical x-ray spectrum are the same. Ideally, separate calibration curves should be generated for different spectra. This has been achieved to a certain extent by collecting the calibration data under automatic exposure control (AEC) but this makes the assumption that AEC performance is temporally stable and that the same x-ray exposure factors (kV, target, filter) would be selected for breasts and tissue equivalent phantoms. Previous work showed no measurable difference in calibration data between a wide range of x-ray beam qualities [37] but this used the PTFE stepwedge ($Z_{\text{eff}} \sim 8.5$) which has an atomic number and attenuation coefficient closer to that of breast tissue.

The next chapter describes the uncertainties in the Manchester Method and quantifies the inaccuracies in volumetric breast density associated with incorrect breast thickness measurement. Studies carried out to validate the current Manchester Method are presented in Chapter 6.
5. Uncertainties in the Manchester Method

This chapter has two components. The first is a discussion of the uncertainties associated with the Manchester Method and their subsequent effect on the accuracy of volumetric breast density measurements. This is followed by an evaluation of intra- and inter-observer variability. The Manchester Method is a semi-automated technique and it is therefore essential that it is shown to be reproducible.

5.1 Sources of uncertainty

A number of uncertainties in the measurements of volumetric breast density are acknowledged and described briefly below. These arise from assumptions associated with the x-ray imaging process and interpolation of calibration data.

5.1.1 Scattered radiation

Attenuation of the x-ray beam has two components: absorption and scatter. Absorption is due to the photoelectric effect, which is proportional to \( \frac{Z^3}{E^3} \). This is the dominant effect in mammography which uses low x-ray beam energy, \( E \), in order to enhance the contrast between fibroglandular and malignant tissues, based on the difference in their atomic number, \( Z \). However, scattered radiation also contributes to the final image. The ratio of scatter to primary radiation is typically about 0.5, although this is highly dependent upon field size and breast thickness, with a lesser dependence on beam energy [49]. However, anti-scatter grids reduce the proportion of scatter reaching the image receptor and experimental work by Dance and Day [183] has shown that approximately 15% of scattered radiation is transmitted for a 6cm breast at 31kV, Mo/Mo.

The tissue equivalent material used to generate the calibration data was designed to simulate the modification of the radiation field caused by absorption and scattering in real breast tissue [142]. For this reason, it is not considered necessary to account for scattered radiation in the Manchester Method, especially as the aim of the technique is determination of breast density, rather than image enhancement.

5.1.2 Anode-heel effect

The anode-heel effect is best described with reference to Figure 5.1 (left). It is a phenomenon whereby x-rays emitted at Point B are attenuated by a greater thickness of anode material than those emitted at Point A, resulting in a non-uniform intensity across the x-ray image. This is demonstrated in Figure 5.1 (right) which is based on the experimental data of Cowen et al [184] when exposing a uniform block of Perspex of thickness 3cm.
In mammography, the x-ray tube is positioned such that the intensity of the x-ray beam decreases from the chest wall to the nipple, which reflects the decrease in breast thickness in this direction. In the current Manchester Method, the decrease in breast thickness is accounted for by the magnification markers on the compression paddle, but the anode-heel effect is not taken into account. This means that glandular tissue volume and percentage breast density will be underestimated, although the extent of this is difficult to quantify as it is dependent upon a number of factors such as beam quality, field size and particularly breast area. The fact that the stepwedge remains in the same location on all images means that there would be no need to apply the anode-heel correction to the stepwedge itself, but ideally it should have been applied over the breast area. However, there will be no need to address this in future work as this will be done on digital mammography images which have already had a “flat-field” correction applied by the manufacturer.

5.1.3 Calibration data

As described in Section 4.4, calibration data were collected by imaging the stepwedge alongside phantoms with varying thicknesses of adipose and glandular tissue equivalent material. This enabled the generation of calibration curves relating stepwedge thickness to glandular tissue thickness for a known total breast thickness. The calibration curves contain a finite number of measured data points and as such, there will be random uncertainties associated with the values of glandular tissue thickness determined from these curves, which will propagate through to the measurements of glandular volume and percentage breast density.

One source of uncertainty is attributed to the process of matching grey levels. For every image, it is necessary to measure the grey level within every step on the stepwedge and use linear interpolation to determine the relationship between stepwedge thickness and grey level. The grey level of every pixel within the phantom (or breast) is then equated to the thickness of stepwedge material with the same grey level.
The slabs of tissue equivalent material have a thickness of 5mm or 10mm. For every total phantom thickness, data were collected for at least three different glandular thicknesses, although five or more thicknesses were usually included. Linear interpolation was again performed between data points, which generates another source of uncertainty in the data.

Systematic errors may be introduced by the fact that the phantoms used for calibration do not have identical attenuation properties to tissue in vivo.

The greatest source of uncertainty in volumetric breast density measurements arises from inaccurate measurements of compressed breast thickness. This is investigated in the following section.

5.2 Effect of inaccurate breast thickness on percentage breast density

The following investigation was carried out in order to quantify the uncertainty in percentage breast density arising from inaccuracies in breast thickness measurement. Normally, the edge of each marker is determined automatically by a Canny edge detector [180], or based on a circle fitted to a number of edge points marked by the operator, using the Hough transform [181]. For each pair of magnification markers, compressed breast thickness is calculated using the distance between the centres of these fitted circles. In order to assess the effect of inaccuracies in breast thickness, the operator deliberately marked points which would cause the centre of the circle to lie on the inner or outer edge of the magnification marker. This had the effect of underestimating or overestimating breast thickness respectively. The error in breast thickness estimation was approximately ±10mm, although this was dependent upon the image.

This analysis was carried out on 10 images with a range of breast thicknesses and densities. In all cases, an underestimate of breast thickness would result in a significant overestimate of glandular volume and percentage breast density whereas an overestimate of breast thickness would have the opposite effect. This is demonstrated in Figure 5.2 which shows the original mammogram (a) and glandular thickness map (b) and the resulting glandular thickness maps for lower (c) and higher (d) breast thicknesses.

Table 5.1 shows results for three images from this sample. These were classed as fatty (density 3.0%, thickness 62.4mm), average (density 14.5%, thickness 33.2mm) and dense (density 28.5%, thickness 57.3mm). The relationship between breast density and breast thickness for each image was used to estimate the change in percentage breast density associated with a ±1mm change in breast thickness, to enable comparison with results in the literature, and a ±5mm change in breast thickness as this represents the tolerance level on system-indicated breast thickness [164].
The results in Table 5.1 are percentage breast density expressed as the absolute value (shaded rows) and the absolute and percentage differences from the original value (unshaded rows).

<table>
<thead>
<tr>
<th>Change in breast thickness relative to original</th>
<th>-5 mm</th>
<th>-1 mm</th>
<th>0</th>
<th>+1 mm</th>
<th>+5 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fatty</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage breast density (%)</td>
<td>9.3</td>
<td>4.3</td>
<td>3.0</td>
<td>2.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Absolute difference from original (%)</td>
<td>6.3</td>
<td>1.3</td>
<td>-</td>
<td>-0.2</td>
<td>-1.2</td>
</tr>
<tr>
<td>Percentage difference from original (%)</td>
<td>209.2</td>
<td>41.8</td>
<td>-</td>
<td>-7.7</td>
<td>-38.6</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage breast density (%)</td>
<td>34.9</td>
<td>18.6</td>
<td>14.5</td>
<td>13.3</td>
<td>8.6</td>
</tr>
<tr>
<td>Absolute difference from original (%)</td>
<td>20.4</td>
<td>4.1</td>
<td>-</td>
<td>-1.2</td>
<td>-5.9</td>
</tr>
<tr>
<td>Percentage difference from original (%)</td>
<td>140.9</td>
<td>28.2</td>
<td>-</td>
<td>-8.2</td>
<td>-40.9</td>
</tr>
<tr>
<td><strong>Dense</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage breast density (%)</td>
<td>41.8</td>
<td>31.1</td>
<td>28.5</td>
<td>26.7</td>
<td>19.6</td>
</tr>
<tr>
<td>Absolute difference from original (%)</td>
<td>13.4</td>
<td>2.7</td>
<td>-</td>
<td>-1.8</td>
<td>-8.9</td>
</tr>
<tr>
<td>Percentage difference from original (%)</td>
<td>46.9</td>
<td>9.4</td>
<td>-</td>
<td>-6.3</td>
<td>-31.3</td>
</tr>
</tbody>
</table>

**Table 5.1: Effect of errors in breast thickness on percentage breast density**
*Unshaded rows: absolute values; shaded rows: percentage difference from original values*
The sample of 10 images had a mean percentage breast density of 13.76% (standard deviation s.d. 9.38%, range 2.01 – 28.46%). The average absolute difference in density was 2.44% (error in thickness of -1mm) and -0.98% (error of +1mm). The average percentage difference in density was much higher, at 24.02% (error of -1mm) and -7.16% (error of +1mm). It can be seen that underestimating breast thickness results in much greater uncertainties in percentage breast density than overestimating breast thickness.

As a comparison, Malkov et al [157] reported an average error in percentage fibroglandular volume of 7% per 1mm thickness deviation using SXA; this increased with decreasing breast thickness. Augustine et al [147] stated that a 3mm error in breast thickness results in ≥15% error in volumetric breast density using their calibration technique. In both cases, it is unclear whether this is the absolute or percentage error, given that the units of measurement are “%”. The effect of errors in breast thickness on percentage breast density using Volpara™ [154] is shown in Table 5.2 for a breast of 32mm with 12.4% density.

<table>
<thead>
<tr>
<th>Variation in breast thickness</th>
<th>-20%</th>
<th>-10%</th>
<th>0%</th>
<th>+10%</th>
<th>+20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast density (%)</td>
<td>15.1</td>
<td>13.6</td>
<td>12.4%</td>
<td>11.4%</td>
<td>10.6%</td>
</tr>
<tr>
<td>Percentage difference from '0%' (%)</td>
<td>20.8</td>
<td>9.7</td>
<td>-</td>
<td>-8.1</td>
<td>-14.5</td>
</tr>
</tbody>
</table>

*Table 5.2: Effect of errors in breast thickness on the density estimate using Volpara™ [154]*

The Manchester Method therefore appears to be very susceptible to errors in breast thickness measurement, compared to Volpara™ although a similar non-linear relationship is exhibited: underestimating breast thickness results in a greater percentage error in breast density than overestimating breast thickness. However, the results are in line with those reported by Malkov et al [157] suggesting that breast thickness measurement must be extremely accurate for calibration techniques. Relying on the system-indicated value of breast thickness is not a viable option.

The magnification markers should provide a measure of compressed breast thickness which is accurate to within 0.4mm. The associated absolute uncertainties in percentage breast density are +0.98% (error of -0.4mm) and -0.39% (error of +0.4mm). These equate to percentage uncertainties of +9.61% and -2.86% respectively.

Studies on observer variation are presented below. In addition to providing an assessment of reliability, it is thought that these will also give an indication of the magnitude of errors in breast thickness and volumetric breast density measurement associated with the Manchester Method.
5.3 Intra-observer analysis of the Manchester Method

Intra- and inter-observer analyses were carried out using the images from the feasibility study (described in Chapters 7 and 8). All statistical analyses were carried out using IBM SPSS Version 19.

One observer (the author) marked up the mammograms of 50 randomly selected women on four separate occasions. Not all women within this sample had four views available for analysis. In total, 187 images were marked up consisting of 96 CC and 91 MLO views. For each parameter generated by the software, the Intraclass Correlation Coefficient (ICC) and the average coefficient of variation (CV) were calculated for the sample as a whole and separately for the CC and MLO groups. The ICC was 1.00 for every parameter indicating excellent intra-observer variability. Results for the coefficient of variation (CV) are shown below in Table 5.3. A CV of zero reflects perfect agreement.

<table>
<thead>
<tr>
<th>Whole sample (n = 187)</th>
<th>Mean</th>
<th>Average CV (%)</th>
<th>Minimum CV (%)</th>
<th>Maximum CV (%)</th>
<th>Perfect agreement (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast thickness (mm)</td>
<td>55.19</td>
<td>0.085</td>
<td>0.000</td>
<td>0.680</td>
<td>92</td>
</tr>
<tr>
<td>Breast area (cm²)</td>
<td>148.73</td>
<td>0.243</td>
<td>0.000</td>
<td>3.617</td>
<td>94</td>
</tr>
<tr>
<td>Breast volume (cm³)</td>
<td>762.86</td>
<td>0.303</td>
<td>0.000</td>
<td>4.094</td>
<td>48</td>
</tr>
<tr>
<td>Glandular volume (cm³)</td>
<td>62.36</td>
<td>1.467</td>
<td>0.063</td>
<td>8.473</td>
<td>0</td>
</tr>
<tr>
<td>Percentage breast density (%)</td>
<td>11.44</td>
<td>1.421</td>
<td>0.050</td>
<td>8.305</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CC views only (n = 96)</th>
<th>Mean</th>
<th>Average CV (%)</th>
<th>Minimum CV (%)</th>
<th>Maximum CV (%)</th>
<th>Perfect agreement (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast thickness (mm)</td>
<td>54.03</td>
<td>0.084</td>
<td>0.000</td>
<td>0.515</td>
<td>48</td>
</tr>
<tr>
<td>Breast area (cm²)</td>
<td>142.06</td>
<td>0.000</td>
<td>0.000</td>
<td>0.009</td>
<td>94</td>
</tr>
<tr>
<td>Breast volume (cm³)</td>
<td>707.99</td>
<td>0.051</td>
<td>0.000</td>
<td>0.370</td>
<td>48</td>
</tr>
<tr>
<td>Glandular volume (cm³)</td>
<td>64.74</td>
<td>1.361</td>
<td>0.063</td>
<td>8.178</td>
<td>0</td>
</tr>
<tr>
<td>Percentage breast density (%)</td>
<td>12.38</td>
<td>1.393</td>
<td>0.063</td>
<td>8.178</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MLO views only (n = 91)</th>
<th>Mean</th>
<th>Average CV (%)</th>
<th>Minimum CV (%)</th>
<th>Maximum CV (%)</th>
<th>Perfect agreement (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast thickness (mm)</td>
<td>56.42</td>
<td>0.087</td>
<td>0.000</td>
<td>0.680</td>
<td>44</td>
</tr>
<tr>
<td>Breast area (cm²)</td>
<td>155.78</td>
<td>0.499</td>
<td>0.043</td>
<td>3.617</td>
<td>0</td>
</tr>
<tr>
<td>Breast volume (cm³)</td>
<td>820.75</td>
<td>0.570</td>
<td>0.092</td>
<td>4.094</td>
<td>0</td>
</tr>
<tr>
<td>Glandular volume (cm³)</td>
<td>59.85</td>
<td>1.580</td>
<td>0.163</td>
<td>8.473</td>
<td>0</td>
</tr>
<tr>
<td>Percentage breast density (%)</td>
<td>10.46</td>
<td>1.450</td>
<td>0.050</td>
<td>8.305</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.3: Coefficient of Variation (CV) for breast parameters generated by the analysis software
Marking up breast thickness was found to be highly reproducible. The difference between the maximum and minimum breast thickness within the four reads was on average only 0.11mm (range 0 - 1.13mm). Perfect agreement was achieved for 92 images (49.2%), split evenly between the CC and MLO views.

The Manchester Method employs a fully-automated technique to detect the breast edge, as described in Section 4.5.2; however, the observer is still required to mark the pectoral muscle. This is reflected in the results for CVs of breast area, which are substantially different between CC and MLO views. Perfect agreement was achieved for 94 of the CC images (97.9%). For the remaining 2 images, the maximum difference in breast area within the four reads was 0.04cm². Although perfect agreement was not achieved for any of the MLO images, marking the pectoral muscle was found to be highly repeatable, with an average CV for the MLO images of less than 0.50% and an average difference between the maximum and minimum area of only 1.65cm² (range 0.22 – 13.74cm²).

The maximum difference of 13.74cm² is considered an outlier and the corresponding image is shown in Figure 5.3. It can be seen that the pectoral muscle is not clearly defined in this mammogram and three possible examples of operator selection are given. The blue line is the most likely, followed by the red line. The black dotted line is less probable but may have been selected on one occasion, given that the results for breast area were 183.8, 183.7, 170.1 and 180.8cm².

![Figure 5.3: Image with the greatest intra-observer variation in breast area due to poorly defined pectoral muscle (left); possible operator selection of pectoral muscle (right)](image_url)

Breast volume is calculated using breast thickness and area, hence results reflect the observations for these parameters. Perfect achievement was achieved for 48 of the CC images
(50.0%), due to the automatic breast edge detection and the highly reproducible method of marking up breast thickness. Agreement was also good for the MLO images with an average CV of 0.57% and an average difference between the maximum and minimum breast volume of 9.85cm$^3$ (range 1.15 – 84.63cm$^3$). The greatest difference was observed for the image in Figure 5.3 and this is attributed to the variation in breast area.

Results for glandular volume and volumetric breast density exhibit larger variation than all other parameters suggesting that marking up the stepwedge is the task with the greatest operator dependence. Results were comparable for CC and MLO views. A CV greater than 5% is considered unacceptable when assessing intra-observer variability. There were 3 images which had CV greater than 5% for glandular volume and 3 images for percentage breast density. As expected, images with a high variation in glandular volume also exhibited a high variation in volumetric breast density, since the two are directly related. Results for these images are shown in Table 5.4.

<table>
<thead>
<tr>
<th>Image</th>
<th>504LML</th>
<th>515RML</th>
<th>548LML</th>
<th>548RCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean breast thickness (mm)</td>
<td>88.43</td>
<td>73.54</td>
<td>72.81</td>
<td>65.87</td>
</tr>
<tr>
<td>Mean breast volume (cm$^3$)</td>
<td>1434.65</td>
<td>1292.39</td>
<td>1087.49</td>
<td>797.46</td>
</tr>
<tr>
<td>Mean glandular volume (cm$^3$)</td>
<td>62.54</td>
<td>69.52</td>
<td>22.23</td>
<td>33.85</td>
</tr>
<tr>
<td>Mean percentage breast density (%)</td>
<td>4.36</td>
<td>5.38</td>
<td>2.04</td>
<td>4.24</td>
</tr>
<tr>
<td>Std deviation in percentage breast density (%)</td>
<td>0.36</td>
<td>0.27</td>
<td>0.10</td>
<td>0.35</td>
</tr>
<tr>
<td>Maximum difference in percentage density (%)</td>
<td>0.69</td>
<td>0.57</td>
<td>0.23</td>
<td>0.70</td>
</tr>
<tr>
<td>CV glandular volume (%)</td>
<td>8.47</td>
<td>4.88</td>
<td>5.09</td>
<td>8.18</td>
</tr>
<tr>
<td>CV percentage density (%)</td>
<td>8.31</td>
<td>5.02</td>
<td>4.77</td>
<td>8.18</td>
</tr>
</tbody>
</table>

Table 5.4: Images with the greatest observer variation in volumetric breast density

With the exception of 548RCC, all are characterised by a breast thickness and breast volume that are significantly higher than the sample means of 55.2mm and 763cm$^3$ respectively. Furthermore, all have percentage breast density which is significantly lower than the sample mean of 11.4%. Given that the CV is calculated as the standard deviation divided by the mean (expressed as a percentage), the simplest explanation for the high CV is the low value in the denominator. However, the standard deviations in percentage breast density for these cases exceed the sample mean of 0.12 (except 548LML) suggesting that this may not fully explain the variation. An additional explanation is offered with reference to images 548RCC and 504LML (Figure 5.4).
The variation in results for images 548LML and RCC may be attributed to the fact that the stepwedge is partially obscured by the ID label. Analysis should still have been possible as the steps are not completely obscured and, as explained in Section 4.6, a correction is made for the grey level of steps being higher or lower than expected. Although the ROI positions and the grey level values associated with each step can be reviewed, the operator may become complacent in checking these when marking up a large number of images.

504LML has particularly large breast thickness and it is likely that the x-ray exposure would have used high kV and mAs. This is reflected in the dark appearance of the stepwedge, which would have made it difficult for the operator to accurately determine the edges of the steps. This was also the case for 515RML.

Despite the fact that percentage breast density has a greater coefficient of variation than other parameters, this still equates to a low absolute difference. Considering the outliers presented in Table 5.4 above, the difference between the maximum and minimum percentage density within the 4 reads is less than 0.70% in all cases. Taking the sample as a whole, the average difference is only 0.27% (range 0.01 – 1.02%). A difference of this magnitude would be highly unlikely to affect a woman's risk profile.

5.4 Inter-observer analysis of the Manchester Method

Two separate studies were carried out to assess inter-observer agreement in the CC and MLO views. More detailed analysis was performed on the CC data in order to determine whether it was acceptable to use the results from two observers in the feasibility study.
5.4.1 CC view

Observer 1 (medical student) carried out the analysis on all CC views from the feasibility study; Observer 2 (the author) analysed all MLO views plus a subset of 168 randomly selected CC images. In order to determine if it was acceptable to use the results from two observers in the feasibility study, Bland-Altman plots [185] were examined. These show the difference (calculated as Observer 2 – Observer 1) against the mean for each pair of measurements and are included for breast thickness, breast volume, glandular volume and percentage breast density in Figures 5.5 - 5.8. The blue line shows the mean difference and the red lines indicate the 95% limits of agreement. The blue arrows represent outliers which are discussed at the end of this section.

Breast area is determined using a fully automated segmentation algorithm and as such, there was perfect agreement between the two observers in 164 of 168 cases. In the remaining 4 cases, the average difference was only 0.33cm$^3$.

Figure 5.5: Bland-Altman plot for breast thickness
Blue arrows show outliers (discussed below)
Figure 5.6: Bland-Altman plot for breast volume
Blue arrows show outliers (discussed below)

Figure 5.7: Bland-Altman plot for glandular volume
Blue arrows show outliers (discussed below)
Further assessment of inter-observer agreement is shown in Table 5.5. Breast thickness followed a normal distribution so association between the observers was measured using the Pearson correlation coefficient, $r$. All other parameters were not normally distributed and therefore Spearman’s rho ($\rho$), a non-parametric correlation coefficient, was used to measure the association between the two observers. The absolute difference in values is quoted as this gives a better indication of the variation between observers than if the difference is stated as Observer 2 – Observer 1. If the latter was calculated for every pair of measurements, the mean difference would lie close to zero, as shown in Figures 5.5 – 5.8, due to a spread of positive and negative differences.

A number of observations can be made, which reflect those for intra-observer variability. There is negligible observer variation in the mark-up of breast thickness, with perfect agreement in 31% of cases. This is also true for breast volume, which is inevitable for the CC view, given that the method for detecting the breast edge is fully-automated and the volume is dependent upon area and thickness. Once again, glandular volume and percentage breast density are the parameters which exhibit the greatest discrepancy between operators, supporting the hypothesis that marking up the stepwedge carries the greatest operator dependence. However, it must be acknowledged that differences in breast thickness also contribute to this finding.
### Table 5.5: Inter-observer variation for breast parameters generated by the analysis software

<table>
<thead>
<tr>
<th></th>
<th>Mean Ob1</th>
<th>Mean Ob2</th>
<th>Correlation</th>
<th>Mean Difference (Absolute)</th>
<th>Min - Max Difference (Absolute)</th>
<th>Perfect agreement (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast thickness (mm)</td>
<td>56.64</td>
<td>56.62</td>
<td>1.000*</td>
<td>0.13</td>
<td>0.000 - 0.90</td>
<td>52</td>
</tr>
<tr>
<td>Breast area (cm²)</td>
<td>161.79</td>
<td>161.78</td>
<td>1.000*</td>
<td>0.01</td>
<td>0.000 - 0.61</td>
<td>164</td>
</tr>
<tr>
<td>Breast volume (cm³)</td>
<td>834.14</td>
<td>834.02</td>
<td>1.000*</td>
<td>1.08</td>
<td>0.000 - 7.99</td>
<td>52</td>
</tr>
<tr>
<td>Glandular volume (cm³)</td>
<td>70.10</td>
<td>69.55</td>
<td>0.995*</td>
<td>1.82</td>
<td>0.010 - 19.27</td>
<td>0</td>
</tr>
<tr>
<td>Breast density (%)</td>
<td>11.10</td>
<td>11.01</td>
<td>0.997*</td>
<td>0.26</td>
<td>0.003 - 2.03</td>
<td>0</td>
</tr>
</tbody>
</table>

*All correlations significant p < 0.001

The Bland-Altman plots (Figures 5.5 – 5.8) show that for all parameters, the mean difference between observers is very close to zero and there is little spread in the data. The results in Table 5.5 indicate that there is excellent correlation (significant, p < 0.001) between the two observers, with many cases of perfect agreement for breast thickness, area and volume. It therefore seems justified to use the results from two observers in the feasibility study.

There are some outliers in the data. Those marked with blue arrows in Figures 5.5 – 5.8 are summarised in Table 5.6. A review of these images did not reveal any obvious reasons for the discrepancies. Markers were clearly visible and well-defined in all images; they did not lie on the film edges and were not obscured by view markers. Additionally, the stepwedge was not over- or under-exposed so the individual steps could be easily resolved. Image 951LCC has the greatest difference in breast thickness between operators and it is likely that this accounts for the relatively large differences in breast volume, glandular volume and percentage breast density. However, for image 059RCC, the differences in volumetric breast density are solely attributable to discrepancies in the mark-up of the stepwedge, given the near perfect agreement in breast thickness.

### Table 5.6: Images with the greatest discrepancy between observers

<table>
<thead>
<tr>
<th>Image</th>
<th>254RCC</th>
<th>951LCC</th>
<th>457RCC</th>
<th>059RCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in breast thickness (mm)</td>
<td>0.80</td>
<td>-0.90</td>
<td>0.47</td>
<td>-0.01</td>
</tr>
<tr>
<td>Difference in breast volume (cm³)</td>
<td>4.75</td>
<td>-7.99</td>
<td>3.73</td>
<td>-0.65</td>
</tr>
<tr>
<td>Difference in glandular volume (cm³)</td>
<td>-5.36</td>
<td>8.66</td>
<td>-16.34</td>
<td>19.27</td>
</tr>
<tr>
<td>Difference in percentage breast density (%)</td>
<td>-1.00</td>
<td>1.29</td>
<td>-2.03</td>
<td>2.00</td>
</tr>
<tr>
<td>Percentage breast density (%): Observer 1</td>
<td>17.80</td>
<td>14.21</td>
<td>11.94</td>
<td>4.99</td>
</tr>
<tr>
<td>Percentage breast density (%): Observer 2</td>
<td>16.80</td>
<td>15.50</td>
<td>9.91</td>
<td>6.99</td>
</tr>
</tbody>
</table>
Once again, it must be stressed that although these are considered outliers, a difference in percentage density of 2.03% is small in absolute terms and would be unlikely to place a woman in a different risk category. Differences of this magnitude were rare; there were only two cases where the difference in percentage breast density exceeded 2% and a further three cases where the difference exceeded 1%.

5.4.2 MLO view

In order to assess inter-observer variation for the MLO view, 61 images were selected which encompassed a range of breast thicknesses and compositions and included both field sizes. These were marked up by two observers (a medical student and the author). A summary of results is shown in Table 5.7. These are very similar to those in Table 5.5 for the CC view, although inevitably there is a greater difference in the results for breast area because the operator must define the pectoral muscle. This affects the results for breast volume, where the discrepancy is also greater than previously observed. However, all correlations are very strong and significant (p < 0.001), suggesting that inter-observer agreement is also excellent for the MLO view.

Observer variation in breast thickness is again very low, with perfect agreement in 11% of cases. The mean absolute difference between readers is only 0.20mm and the maximum difference is 1.04mm. However, this outlier refers to a large breast (87.07mm and 88.11mm for Observer 1 and 2 respectively) so the markers would have been close to the edges of the film and it may have been difficult to accurately determine their position. This discrepancy in breast thickness had little effect on percentage breast density which was measured as 1.93% and 2.03% for Observers 1 and 2 respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean Ob1</th>
<th>Mean Ob2</th>
<th>Correlation</th>
<th>Average Difference (Absolute)</th>
<th>Min – Max Difference (Absolute)</th>
<th>Perfect agreement (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast thickness (mm)</td>
<td>59.68</td>
<td>59.67</td>
<td>1.00*</td>
<td>0.20</td>
<td>0.000 – 1.04</td>
<td>7</td>
</tr>
<tr>
<td>Breast area (cm²)</td>
<td>168.41</td>
<td>168.62</td>
<td>0.999*</td>
<td>1.13</td>
<td>0.001 – 8.91</td>
<td>0</td>
</tr>
<tr>
<td>Breast volume (cm³)</td>
<td>961.18</td>
<td>963.33</td>
<td>1.00*</td>
<td>7.46</td>
<td>0.20 – 43.35</td>
<td>0</td>
</tr>
<tr>
<td>Glandular volume (cm³)</td>
<td>79.98</td>
<td>81.60</td>
<td>0.996*</td>
<td>2.83</td>
<td>0.003 – 19.35</td>
<td>0</td>
</tr>
<tr>
<td>Breast density (%)</td>
<td>13.24</td>
<td>13.46</td>
<td>0.998*</td>
<td>0.38</td>
<td>0.001 – 1.93</td>
<td>0</td>
</tr>
</tbody>
</table>

* All correlations significant p < 0.001

Table 5.7: Inter-observer variation for breast parameters generated by the analysis software 61 images, MLO view
Glandular volume and percentage breast density are again the parameters which suffer from the greatest observer variability. Image 806LML (Figure 5.9) has the largest difference in glandular volume and percentage breast density (Observer 1: 116.5cm$^3$ and 31.6%; Observer 2: 119.8cm$^3$ and 33.5%). However, reasons for the observer discrepancy are unclear.

Figure 5.9: Image with the greatest observer variation in volumetric breast density, 806LML

5.5 Discussion

The Manchester Method has been found to demonstrate excellent inter- and intra-observer agreement. The process of selecting the magnification markers for breast thickness measurement was found to be highly reproducible, with an average difference in breast thickness of 0.11mm (intra-observer), 0.13mm (inter-observer, CC view) and 0.20mm (inter-observer, MLO view). Percentage breast density and glandular volume were also found to be very reproducible, but to a lesser extent than breast thickness, suggesting that marking up the stepwedge is the task with the greatest operator variability. Average absolute differences in percentage breast density were calculated as 0.27% (intra-observer), 0.26% (inter-observer, CC view) and 0.38% (inter-observer, MLO view).
6. Validation and Comparison

In general terms, validation is defined as the process taken to ensure that a product or technique meets the initial requirements and specifications for which it was developed. In the case of the Manchester Method, this means that the measurements of breast density generated by the technique must exhibit a strong association with breast cancer risk, or with known risk factors. The method must also be capable of identifying women with high breast density as the sensitivity of mammography will be lower in this group and follow-up examinations using other imaging modalities may be appropriate.

One of the aims of the feasibility study was to examine correlations between volumetric breast density (glandular volume and percentage breast density) and well-established breast cancer risk factors; this is therefore considered part of the validation process and results are presented in Chapter 8.

The relationship between breast density and breast cancer risk is perhaps of greater interest. This would best be determined using a case-control study which includes the priors of screen-detected and interval-detected cancers. However, the Manchester Method has not yet been applied in this way. The feasibility study was a prospective study on the screening population and as such, the number of cancers within the study cohort will be small. Assessing the relationship between volumetric density and breast cancer risk will be possible, but limited.

An alternative strategy would be to compare the results of the Manchester Method to those from a technique which has shown a strong and well-established correlation with breast cancer risk. This is the approach described in this chapter. Cumulus is considered to be the gold standard for estimating mammographic density and has been used in studies of this nature [118, 119]. However, visually assessed mammographic density is also strongly associated with breast cancer risk [18, 21, 116] so the measurements of volumetric density using the Manchester Method are compared to mammographic density assessed by two radiologists.

The Manchester Method can be adapted to measure percentage breast density by area. The method generates the thickness of gland and fat within each column of tissue projected onto a pixel in the digitised image. Expressing the glandular thickness as a percentage of the total breast thickness within each pixel allows the pixel to be defined as fatty or dense, based on a specified threshold of glandularity. These area-based measurements are also compared to visually assessed mammographic density; this should provide an insight into how radiologists classify a dense pixel.
The data presented in this chapter are based on analysis of images from the feasibility study (described in Chapters 7 and 8). All statistical analyses were carried out using IBM SPSS Version19.

6.1 Inter-observer analysis of visually-assessed mammographic density

In Chapter 5, the Manchester Method was shown to demonstrate excellent intra- and inter-observer agreement. The results are compared to the inter-observer variation associated with visually assessed mammographic density.

6.1.1 Method

Two radiologists used a Visual Analogue Scale (Figure 3.3) to estimate density on the mammograms of 167 randomly selected women (668 images). Both readers were consultant radiologists and therefore highly experienced in mammographic interpretation. Radiologist 2 had greater experience than Radiologist 1 in using visual analogue scales for density estimation, having acted as one of the eight radiologists in the study by Duffy et al [116], where mammographic density was assessed for 10,048 women.

The results for each radiologist were initially categorised using the Boyd Six Category Classification (SCC) scheme [16]. Some variation between radiologists was expected, but if the mammographic density differed by more than two categories, the case was removed. The reason for doing this was that Radiologist 1 read the images for a study in 2007 [186] and Radiologist 2 read the images for a study in 2009 [187], hence the most likely explanation for a significant variation in density was that Radiologist 2 had not been provided with an identical sample of images. On this basis, 5 women were removed from the sample, leaving a sample size of 162 women (648 images).

6.1.2 Results

The following analysis treats mammographic density as a continuous variable. Paired-sample t-tests revealed that for both radiologists, there was no significant difference between the mammographic density in the left and right breasts (using the average value of the CC and MLO view). Therefore, graphs are presented for the left breast only.

Histograms of percentage mammographic density for both radiologists are shown in Figure 6.1. It can be seen that the distributions are quite different. The results for Radiologist 1 are predominantly skewed to the left although there is a small peak at high density (around 70%). The distribution for Radiologist 2 is more uniform although there appear to be three distinct peaks at low (10%), mid (40%) and high (70%) densities.
Figure 6.1: Histograms of percentage mammographic density for two radiologists assessing a sample of 324 left breast images; the average of the LCC and LMLO densities are shown

Descriptive statistics for the average mammographic density of the left and right breasts, for each radiologist, are shown in Table 6.1. Paired t-tests were used to analyse the differences between the two radiologists and results are shown in Table 6.2. The mean difference in mammographic density was substantial and significant (p < 0.001) for both breasts, with Radiologist 2 consistently estimating higher mammographic densities than Radiologist 1.

<table>
<thead>
<tr>
<th>n = 162</th>
<th>Mean (%)</th>
<th>Std Deviation (%)</th>
<th>Median (%)</th>
<th>Minimum (%)</th>
<th>Maximum (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Left</td>
<td>Right</td>
<td>Left</td>
<td>Right</td>
</tr>
<tr>
<td>Radiologist 1</td>
<td>27.38</td>
<td>23.80</td>
<td>18.75</td>
<td>1.50</td>
<td>89.25</td>
</tr>
<tr>
<td>Radiologist 2</td>
<td>43.91</td>
<td>25.71</td>
<td>39.75</td>
<td>4.00</td>
<td>93.00</td>
</tr>
</tbody>
</table>

Table 6.1: Visually assessed mammographic density

<table>
<thead>
<tr>
<th>Radiologist 2 – Radiologist 1: Mammographic density (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Difference</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Left breast</td>
</tr>
<tr>
<td>Right breast</td>
</tr>
</tbody>
</table>

Table 6.2: Variation in mammographic density between radiologists

The reliability of the data was assessed using the average measures Intraclass Correlation Coefficient (ICC). When testing for absolute agreement between average left breast density, the ICC was 0.830 (95% CI: -0.014 – 0.943). This wide confidence interval suggests very poor levels of agreement for some images.
However, when testing for consistency, the ICC was 0.929 (95% CI: 0.903 – 0.948) which indicates that despite the differences in their absolute estimates of mammographic density, the radiologists classify different levels of density in the same way. This is illustrated more clearly on the Bland-Altman plot for the left breast (Figure 6.2) which shows the difference (calculated as Radiologist 2 – Radiologist 1) against the mean for each pair of measurements.

![Bland-Altman plot](image)

*Figure 6.2: Bland-Altman plot for visually-assessed mammographic density by two radiologists*

It can be seen that the data points are clustered around a difference of zero at the extremes of density, indicating that inter-observer agreement is strongest when estimating densities of very fatty or very glandular breasts. The least agreement was found for heterogeneous breasts, which fall in the middle of the density range.

Because mammographic density is not normally distributed (Figure 6.1), the Spearman correlation coefficient (\( \rho \)) was used to test for statistical dependence. The correlation between the two radiologists was found to be statistically significant (\( p < 0.001 \)) for both the left (\( \rho = 0.920 \)) and right breasts (\( \rho = 0.918 \)).

### 6.1.3 Discussion of findings

It is concluded that, in this study, mammographic density suffers from inter-observer variability when measured as a continuous variable using a visual analogue scale. Overall, absolute agreement was poor, as shown by the wide confidence interval on the ICC for agreement and the mean difference in density assessment between radiologists of 15.7%. However, good
agreement was observed for very low and very high breast densities (Figure 6.2). Similar results were reported by Ciatto et al [121] who found that, overall, moderate agreement was achieved between radiologists using the BI-RADS categories [28], but major disagreement was observed in allocating BI-RADS Categories 2 and 3 (25 – 50% and 50% – 75% density).

It is reassuring to find that the ICC for consistency was high (0.929) and the results of the radiologists were well-correlated ($\rho = 0.920, p < 0.001$) which suggests that radiologists classify mammographic density in a consistent fashion.

The fact that the radiologists read the images on different occasions is a limitation of the study. There may have been an error in providing the exact same set of images, although extreme outliers in data were removed. Furthermore, Radiologist 1 assessed density on the original film-screen mammograms but Radiologist 2 used the digitised images. However, prior to commencing the study, she read a sample of 34 original film images (17 CC, 17 MLO) and mammographic density was not significantly different between film and digitised images (mean difference: 1.06%, 95% CI: -1.390 to 3.508, $p = 0.373$).

An additional limitation is that density assessment using the visual analogue scale is not routine practice in the UK. It is thought that much better inter-observer agreement would have been achieved if the radiologists had been trained in using the visual analogue scale with a reference set of images. Radiologist 1 commented on the difficulties of using the scale, particularly when evaluating images with many small areas of scattered glandularity. Furthermore, he stated that he would have found it easier to express the percentage density as a number, which he tended to do mentally before marking it on the scale.

The subjectivity of mammographic density estimation using the visual analogue scale is a clear disadvantage. The inter-observer agreement demonstrated by the Manchester Method is far superior to that of visual assessment, based on this study.

However, the main advantage of the technique is the speed. Radiologist 2 assessed 668 digitised images in less than two hours, which included the time taken for the computer to load and display each image. If density was routinely estimated at the time of reading the mammogram, it would add less than 10s to the reading time for each image. In Section 4.6, the analysis time of the Manchester Method was quoted as 55s per image. Radiologist-assessed mammographic density has also shown a strong association with breast cancer risk and for this reason, the relationship between volumetric breast density and mammographic density is examined in the next section.
6.2 Correlation of volumetric breast density using the Manchester Method with visually assessed mammographic density

The box and whisker plot in Figure 6.3 shows the spread of breast density in the LCC view. Data are presented for mammographic density estimated by both radiologists and volumetric percentage breast density measured using the Manchester Method. The dark line in the middle of the boxes is the median; the bottom and top of the boxes indicate the 25th and 75th percentiles respectively. The whiskers extend to 1.5 times the height of the box or, if no case has a value in that range, to the minimum or maximum values. The circular points are outliers.

It is immediately apparent that volumetric breast density is much lower than mammographic density; this is expected. On a digitised or digital mammogram, a pixel which is classed as dense will not be composed solely of glandular tissue. In fact, it is commonly assumed that the breast has a superficial 5mm adipose layer all the way around [167]. The composition of a ‘dense’ pixel is examined in Section 6.3.

It is therefore not possible to assess the agreement between mammographic density and volumetric breast density. Instead, the correlation between area and volume-based measurements is examined using Spearman’s $\rho$, as the data follows a non-parametric distribution. Results are shown in Table 6.3 for all views. All correlations were significant ($p < 0.001$).
Correlation ($\rho$) of visually assessed mammographic density with:

<table>
<thead>
<tr>
<th>View</th>
<th>Percentage breast density (%)</th>
<th>Glandular volume (cm$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Radiologist 1</td>
<td>Radiologist 2</td>
</tr>
<tr>
<td>LCC</td>
<td>0.750</td>
<td>0.703</td>
</tr>
<tr>
<td>LMLO</td>
<td>0.706</td>
<td>0.731</td>
</tr>
<tr>
<td>Left breast</td>
<td>0.741</td>
<td>0.736</td>
</tr>
<tr>
<td>RCC</td>
<td>0.727</td>
<td>0.743</td>
</tr>
<tr>
<td>RMLO</td>
<td>0.721</td>
<td>0.759</td>
</tr>
<tr>
<td>Right breast</td>
<td>0.738</td>
<td>0.766</td>
</tr>
</tbody>
</table>

**Table 6.3: Correlation of mammographic density with volumetric breast density**

The correlation with mammographic density is stronger for percentage breast density than glandular volume. This was also observed by Jeffreys et al [119] when comparing breast density measured by Volpara™ [154] to percentage mammographic density measured on raw digital mammography images by Cumulus v4. There was strong correlation between percentage breast density and Cumulus ($\rho = 0.85$) but this was much lower for absolute glandular volume ($\rho = 0.45$). Kontos et al [118] compared percentage breast density using Quantra™ [153] to percentage mammographic density measured on processed digital mammography images by Cumulus v4 and found them to be strongly correlated (Pearson’s $r = 0.79$, $p < 0.001$).

Although there was poor absolute agreement in mammographic density between the two radiologists, it is interesting to note that this is not reflected in the strength of correlation with volumetric breast density, which is approximately the same for both radiologists. The analysis presented below uses only the mammographic density assessed by Radiologist 2.

Linear and non-linear regression analysis was carried out to determine the relationship between volumetric breast density measured by the Manchester Method and mammographic density, using data from the left breast (average of CC and MLO views). Equations took the form $bx + c$ (linear) or $ax^2 + bx + c$ (quadratic), where $c$ was selected to be a constant or take the value of zero. Forcing the line of fit to go through the origin resulted in a much higher strength of fit ($R^2$). Figures 6.4 and 6.5 show the relationship between percentage breast density and glandular volume respectively with mammographic density. The linear and non-linear equations are included for the lines of fit which pass through the origin.
Figure 6.4: Relationship between percentage volumetric breast density and mammographic density for the left breast. Linear and quadratic lines of fit are displayed.

Figure 6.5: Relationship between glandular volume and mammographic density for the left breast. Linear and quadratic lines of fit are displayed.
All relationships were found to be statistically significant ($p < 0.001$). The strength of fit was greater for percentage breast density than glandular volume, which is expected based on the results in Table 6.3. For percentage breast density and glandular volume, the quadratic fit ($R^2 = 0.824$, $R^2 = 0.754$ respectively) was slightly stronger than the linear fit ($R^2 = 0.816$, $R^2 = 0.752$ respectively). Similar observations were made by Kontos et al [118] when comparing volumetric density from Quantra™ to mammographic density using Cumulus v4. Statistically significant associations ($p < 0.001$) were found for both models but the non-linear second-degree polynomial had a stronger fit ($R^2 = 0.70$) compared to the linear fit ($R^2 = 0.62$).

Percentage breast density using the Manchester Method was also compared to the BI-RADS categories [28]. Images were allocated to one of the four categories (<25%, 25 - 50%, 50 - 75% and >75%) based on the mammographic density estimated by Radiologist 2. The corresponding value of percentage breast density for the images within each category is shown in Figure 6.6. The dark line in each box is the median value and the box contains the interquartile range.

![Figure 6.6: Association between percentage breast density and the BI-RADS categories. Classification is based on mammographic density assessment by Radiologist 2](image)

Visually, there is a strong relationship between percentage breast density measured by the Manchester Method and the BI-RADS categories. However, this is non-linear and there is some overlap in the percentage breast density of women assigned to each category. A similar association was observed between Volpara™ and the BI-RADS categories. The median breast density in the lowest category was 4.0% and rose to 18.9% in the top category [119].
The linear and quadratic equations for the lines of fit shown in Figure 6.4 were used to calculate the thresholds of percentage breast density which would correspond to the BI-RADS thresholds of 25%, 50% and 75% mammographic density. These are shown in Table 6.4 below, alongside the number of women and the median percentage breast density within each category.

<table>
<thead>
<tr>
<th>Category based on:</th>
<th>Category</th>
<th>Number in category</th>
<th>Median percentage density in category (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammographic density (BI-RADS)</td>
<td>1. &lt; 25%</td>
<td>42</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>2. 25 – 50%</td>
<td>53</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>3. 50 – 75%</td>
<td>37</td>
<td>12.4</td>
</tr>
<tr>
<td></td>
<td>4. &gt;75%</td>
<td>25</td>
<td>23.1</td>
</tr>
<tr>
<td>Linear equation</td>
<td>1. &lt; 6.4%</td>
<td>54</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>2. 6.4 – 12.8%</td>
<td>56</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>3. 12.8 – 19.1%</td>
<td>17</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td>4. &gt; 19.1%</td>
<td>30</td>
<td>25.0</td>
</tr>
<tr>
<td>Quadratic equation</td>
<td>1. &lt; 4.6%</td>
<td>32</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>2. 4.6 – 10.4%</td>
<td>61</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>3. 10.4 – 17.5%</td>
<td>31</td>
<td>12.4</td>
</tr>
<tr>
<td></td>
<td>4. &gt; 17.5%</td>
<td>33</td>
<td>24.3</td>
</tr>
</tbody>
</table>

Table 6.4: Comparison of the distribution of percentage breast density data within the existing BI-RADS categories and the proposed categories based on linear and quadratic equations

If the linear equation was used to determine the thresholds for the categories of percentage breast density, a substantial proportion of women would be allocated to the low risk categories (1 and 2) and only 47 would be allocated to the high risk categories (3 and 4). However, if the quadratic equation was used to determine the thresholds for the categories, 64 women would be assigned to the high risk categories, compared to 62 using the existing thresholds. This approach has potential in studies investigating the association of volumetric breast density with breast cancer risk.

Visually assessed mammographic density has been found to be strongly associated with breast cancer risk [116]. The results in this section show that percentage breast density measured using the Manchester Method exhibits strong and significant (p < 0.001) correlations with visually assessed mammographic density when this variable is treated as continuous and categorical (using the BI-RADS classification). This is also true for glandular volume although the strength of correlation is lower. These are promising findings which suggest that volumetric breast density should also be related to breast cancer risk.
6.3 Correlation of area-based breast density using the Manchester Method with visually assessed mammographic density

In order to gain an insight into the thickness of glandular tissue corresponding to ‘dense’ pixels as defined by visual assessment, the Manchester Method has been adapted to compute an area-based measure of breast density.

The method provides an estimate of the total breast and glandular tissue thicknesses within every column of tissue projected onto a pixel in the digitised image. It was developed as a volumetric technique so ordinarily these would be summed over the entire breast area to compute the total breast volume, the glandular volume and consequently volumetric percentage breast density. However, by expressing the thickness of glandular tissue as a percentage of the total breast thickness at any point, thresholds can be assigned to classify the pixel as ‘dense’. The number of dense pixels is summed and expressed as a percentage of the number of pixels within the breast, thereby providing an area-based measure which can be compared to mammographic density. The aim of this study was to simulate the visual assessment process and determine which threshold corresponds most closely to the way in which radiologists classify dense pixels.

6.3.1 Method

Six thresholds were assigned for classifying a dense pixel; these were greater than 0, 5, 10, 15, 20 and 25% glandularity. As an example, consider a pixel where the grey level corresponds to a composition of 12% glandular tissue. At the 15% threshold, this pixel would be treated as non-dense tissue (fatty background) whilst at the 10% threshold, it would be classified as dense tissue. At the 0% threshold, any pixel which contained even the smallest amount of gland was classed as dense.

These are shown in Figure 6.7 alongside an original mammogram image (CC view). The binary glandular tissue maps for each threshold are shown in Figure 6.8 for this mammogram image. At a threshold of 0%, virtually the whole breast appears to be dense. As the threshold increases, the area of dense tissue decreases. The percentage area of dense tissue generated by the software at each of the six thresholds was compared with the visual assessment of Radiologists 1 and 2.

6.3.2 Results

Results were similar regardless of mammographic view so are presented for the LCC view only in Table 6.5. Correlations between the radiologists’ estimates and the software-generated values of percentage density were assessed using Spearman’s ρ, due to the non-parametric nature of the data.
Figure 6.7: Original mammogram image and the six thresholds for classifying a dense pixel.

Figure 6.8: Binary glandular tissue maps at each threshold for the image shown in Figure 6.7. Dense glandular regions are shown as white.
Table 6.5: Mean percentage dense area generated by the Manchester Method at each threshold and correlations with radiologist assessed mammographic density

As expected, at thresholds of 0% and 5%, the software produced values that were consistently far higher than the radiologists’ estimates. At the 25% threshold, the software values were much lower than the radiologists’ estimates, particularly for Radiologist 2. The strongest agreement was exhibited for the 10, 15 and 20% thresholds and these relationships were investigated further using Bland-Altman plots shown in Figures 6.9 – 6.11. The difference has been calculated as Software – Radiologist, the blue line shows the mean difference and the red lines indicate the 95% limits of agreement.

Radiologist 2 exhibited the best agreement with the 10% threshold, indicated by a mean difference close to zero with a high proportion of data points lying close to this line and an approximately equal spread of data above and below this line. Paired sample t-tests showed that this was the only threshold for which there was no significant difference between the software-generated values and the mammographic density assessed by Radiologist 2 (mean difference: -0.09%, 95% CI: -3.77 – 3.60%, p = 0.964). The agreement between Radiologist 2 and the Manchester Method for dense area became worse as the threshold was increased, demonstrated by a negative mean difference and the distribution of data points, which were almost all below the line of zero difference.

The way in which Radiologist 1 defined a dense pixel was most closely represented by the 15% threshold, although reasonable agreement was also demonstrated visually with the 20% threshold. However, paired sample t-tests showed that 15% was the only threshold for which there was no significant difference between the software-generated values and the mammographic density assessed by Radiologist 1 (mean difference: 2.24%, 95% CI: -0.79 – 5.27%, p = 0.146). The measurements of dense area generated by the Manchester Method using the 10% threshold were typically higher than those of Radiologist 1, as illustrated by the high proportion of data points above the line of zero difference.
Figure 6.9: Bland-Altman plots for percentage dense area using Manchester Method with 10% threshold and radiologist visual assessment (left: Radiologist 1; right: Radiologist 2)

Figure 6.10: Bland-Altman plots for percentage dense area using Manchester Method with 15% threshold and radiologist visual assessment (left: Radiologist 1; right: Radiologist 2)

Figure 6.11: Bland-Altman plots for percentage dense area using Manchester Method with 20% threshold and radiologist visual assessment (left: Radiologist 1; right: Radiologist 2)
A common trend for the thresholds shown in Figures 6.9 – 6.11 is that agreement between both radiologists and the Manchester Method is much greater at the extremes of percentage dense area than for breasts which fall in the intermediate range of density. A cluster of points around the line of zero difference is observed for breasts that are very fatty or very glandular. The same finding was reported for the inter-observer comparison between the two radiologists (Figure 6.2).

6.3.3 Discussion of findings

When visually assessing mammographic density, the radiologist defines a dense pixel as one in which the percentage of glandular tissue is at least 10 to 20% of the total thickness of the compressed breast at that point. This finding is reflected in the images in Figure 6.8. Given that Radiologist 2 estimated significantly higher mammographic density than Radiologist 1, it is unsurprising that their assessment of mammographic density exhibited better agreement with the software-generated values using a lower threshold.

The results provide further evidence that mammographic density is not a true reflection of breast density although the relationship between the two variables is not straightforward. It is certainly not possible to use the threshold value to convert percentage dense area to percentage dense volume, as the threshold represents the minimum percentage of glandular tissue within a column of breast tissue and there will be pixels where the percentage glandularity is much higher than this. There will also be pixels where there is a glandular component but if this were to lie below the threshold value, it would not be included. Additionally, the range of compressed thickness across the breast means that 10% of glandular tissue close to the nipple represents a much lower absolute thickness of gland than 10% of glandular tissue close to the chest wall. An alternative strategy would be to define thresholds in terms of absolute thickness of glandular tissue, rather than percentage thickness. This certainly warrants investigation in future studies although it is less clearly related to radiologists’ perception of density.

Because no single threshold demonstrated perfect agreement with either radiologist, it is not proposed that the Manchester Method be used in this way to provide semi-automated measures of percentage dense area. However, the application of thresholds could be used to aid with the determination of tissue composition within the breast margin. Figure 6.8 indicates that the Manchester Method overestimates the volume of dense tissue in this region; this is particularly noticeable at the lower thresholds (5 - 10%) where there is a thick white band at the breast edge. The results from Chapter 5 suggest that this is likely to result from underestimating the breast thickness within the margin. Because the breast margin is not in contact with the compression plate, the thickness must be modelled in this region. This is achieved using an elliptical profile, as described in Section 4.5.2. As an alternative to excluding the entire breast margin from volumetric breast density analysis, a pragmatic approach would be to use a low percentage threshold of glandular tissue as the basis for exclusion.
6.4 Discussion

The excellent intra- and inter-observer agreement demonstrated by the Manchester Method (Chapter 5) is far superior to that exhibited by two radiologists using the visual analogue scale to estimate mammographic density. However, better inter-radiologist agreement would have been achieved if they had been trained in using the visual analogue scale with a reference set of images.

Percentage breast density and glandular volume generated by the Manchester Method were found to be highly correlated with mammographic density when this was treated as both a continuous and categorical variable; measurements were categorised using the BI-RADS classification [28]. The strong and significant correlations between percentage breast density and mammographic density suggest that volumetric breast density may also be related to breast cancer risk. This is investigated further in Chapter 8 using data from the feasibility study.

It is interesting to note that validation studies on breast density measurement techniques do not generally include an evaluation of technique accuracy; this is particularly true for area-based techniques, as it is acknowledged that mammographic density is not a measure of glandular tissue volume. The accuracy of some volumetric breast density methods has been assessed using tissue-equivalent phantoms [145, 153, 154, 159], or by comparison with truly three-dimensional (3D) techniques, such as magnetic resonance imaging (MRI) [153, 154] or dedicated breast computed tomography (DBCT) [158, 159]. However, there are still uncertainties associated with measurements made using these 3D systems. Studies have reported that dense tissue is hand-segmented on MRI slices of 5mm thickness using the PACS workstation [153, 154] which will inevitably lead to inaccuracies.

The accuracy of volumetric breast density measurements using the Manchester Method has not yet been compared to equivalent measurements made using 3D techniques. This represents an area of future work. The validity of using tissue equivalent phantoms to evaluate the accuracy of the method seems questionable, given that the calibration data were derived using such phantoms. Although it would be possible to compare density measurements made using the Manchester Method to those from other techniques reported in the literature, this is limited by confounding factors within the study populations such as age and weight.
7. Application of the Manchester Method to the Screening Population

In 2007, ethics approval was granted by the Bolton Local Research Ethics Committee to carry out a feasibility study on the use of the Manchester Method of breast density measurement in the UK National Health Service Breast Screening Programme (NHSBSP). Funding was provided by John Lewis plc and Genesis, the breast cancer prevention charity. The study title was “Breast density measurement derived from analysis of screening mammograms for improving breast cancer risk assessment.”

This was the first time the Manchester Method had been applied to the screening population. The study provided an opportunity to assess the feasibility of using the new, compact stepwedge in a clinical environment.

The Bolton, Bury and Rochdale Breast Screening Programme acted as the participating centre and the support of Dr Tony Maxwell (Clinical Director) and Claire Mercer (Superintendent Radiographer) is gratefully acknowledged. The study took place at two of their screening sites: the static site at Bury (Popham Centre) and a mobile van (located at Pike’s Lane Health Centre, Bolton). It started on 9 July 2007 and data collection was completed in December 2007. Digitisation and analysis were carried out over the following 2 years.

This chapter describes the study methodology and provides descriptive statistics on the distribution of volumetric breast density in the screening population. Chapter 8 examines the relationship between breast density and risk-related information collected by questionnaire during the study.

7.1 Aims of the study

The two main aims of the study were to:

1. Determine the practicality of using the Manchester Method, with the new stepwedge, for measuring breast density from mammograms in the screening programme. The new stepwedge refers to the compact aluminium model with the improved attachment mechanism.
2. Ascertain whether additional information, relevant to individual breast cancer risk, can be obtained from women attending for routine screening, using a questionnaire.

There are several aspects involved in assessing how practical the method is. One aspect is purely to determine if it is convenient to use in a screening environment where appointment times are typically 6 minutes. During this time, the identity of the woman is confirmed, the woman gets
changed and two exposures of each breast are made in both the cranio-caudal (CC) and medio-lateral oblique (MLO) views. The women attending are likely to be anxious and will expect a certain level of personal attention and reassurance from the radiographer carrying out the examination. The radiographer will be focused on fulfilling the clinical requirements of the examination and it is vital that the method does not impede on this. However, in order to be successful, it is necessary for the radiographer to ensure that the stepwedge and marker sheets are included in the correct location on each breast image.

In addition to the practicalities, it is necessary to determine the adequacy of the method in measuring the breast density of women in the breast screening programme, where breasts will vary in thickness from approximately 2cm to 10cm and from predominantly fatty to predominantly glandular in composition.

The reasons for collecting risk-related information via questionnaire are discussed in detail in Chapter 8. Analysing correlations between well-established risk factors and breast density may provide a better understanding of breast density variation within the screening population. Additionally, data collected in studies of this nature could be used to develop risk prediction models.

7.2 Method

6,000 women invited for routine breast screening in the Bolton, Bury and Rochdale Breast Screening Programme were invited to participate in the study. An information sheet, covering letter and brief questionnaire were sent to each woman prior to her appointment. This documentation is included as an Appendix to the thesis. Invitations for screening appointment are sent out three weeks before the appointment. The study information was sent separately two days after the invitation, the main reason being that only two sheets of paper can be included per envelope when using an automatic folding machine. However, it was also thought that this would improve the chances of the study documentation being read as women may only read the first page of the documentation, which would be the screening invite, and not realise that they had also been invited to participate in a study. My work telephone number and mobile phone number were provided on the information sheet for women to contact me if they had any queries associated with the study. A log of the queries was kept.

The questionnaire gathered information on date of birth, height and weight (to enable calculation of body mass index (BMI)), date of first pregnancy, ages of menarche and menopause, ethnicity and family history of breast cancer (mother or sister only) including the age at which breast cancer was diagnosed in this relative. For each question, the woman was asked to tick a box to state how certain she was about her answer. There were four responses in this Likert Scale [188]. These were ‘don’t know’, ‘not sure’, ‘quite sure’ and ‘certain.’ It was hoped that this would provide an
approximate indication of the reliability of the data. In order to quantify the error associated with the weight data provided, it was anticipated that a sample of the women attending at the Bury site would be weighed using calibrated scales and the actual weight recorded alongside the estimate. Further relevant information including use of hormone replacement therapy (HRT) and details of present and previous symptoms is routinely recorded in the patient’s notes.

Women wishing to take part in the study were asked to take their completed questionnaire with them when they attended for screening. Informed consent was obtained by the radiographer at Bolton, or the receptionist at Bury. Consent forms were not sent to the woman but were kept at the screening sites. In order to facilitate convenient completion of the consent forms, date stamps were provided and the two receptionists at Bury were also provided with a name stamp. All clerical and radiographic staff involved with the study were provided with an instructional leaflet and several visits were made to each site to demonstrate how to position the stepwedges (one for each field size) and marker sheets (one for each compression paddle) and to address any questions or concerns.

Mammograms were taken as usual. The stepwedge was clipped to the breast support platform alongside the view markers (Figure 7.1a). The only time it would be moved out of the way (Figure 7.1b) was if it was too tall to fit under the compression paddle alongside a very small breast (compressed breast thickness $\leq 14$mm), or there was too little space for it to fit alongside an extremely large breast. In these cases, the radiographer was asked to note this on the consent form. The magnification marker sheets remained on the compression paddle at all times. Neither the stepwedge nor the markers would come into contact with the breast and previous work has demonstrated that they should not interfere with reading the mammogram [182].

Following film reading by the radiologists, mammograms were digitised using the Vidar CadPro digitiser and anonymised. All digitised images were analysed using the Manchester Method; full details of these processes were given in Chapter 4.
Questionnaire data and breast composition data for every woman participating in the study were recorded in a specially developed database created using Microsoft Excel.

7.3 Results

The outputs of the breast density software are compressed breast thickness (maximum value as measured at the chest wall), breast area, breast volume, absolute glandular tissue volume and percentage breast density (calculated as glandular volume / breast volume x 100%). Descriptive statistics on these parameters and a discussion of outliers and method failures are included in the following sections. Analysis of questionnaire data and the correlation of breast density with risk factors are presented in Chapter 8. All statistical analyses were carried out using IBM SPSS Version 19.

7.3.1 Consent rate

Of the 6,000 women invited to participate in the study, 1,414 women completed a consent form, giving a consent rate of 23.6%. Of these, 1,401 responded to the questionnaire. A number of films were not digitised as breast density analysis would have been impossible on these images due to a missing stepwedge or markers (see Section 7.3.5 for full discussion). There were 1,289 women in the study with a completed questionnaire and breast density data for at least one view. The effective participation rate of the study is therefore 21.5%. For all women included in the study, 75% (n = 964) attended at Bury and 25% (n = 325) attended the mobile van at Bolton. Assuming an equal number were invited to Bury and Bolton, the site participation rates were 32.1% and 10.8% respectively.

7.3.2 Descriptive Statistics

This section contains statistics on the average, range and distribution of data for compressed breast thickness, area and volume, glandular tissue volume and percentage breast density.

7.3.2.1 Compressed breast thickness

There were 880 women with data for all 4 views; compressed breast thickness was calculated for each woman as the mean of all 4 views. Within this sample, the mean compressed breast thickness was found to be 59.0mm (standard deviation, s.d. 11.4mm). The breakdown by view is shown in Table 7.1.

As expected, the compressed breast thickness in the MLO view, which contains pectoral muscle, is typically greater than that in the CC view. The average values of breast thickness in the Bolton and Bury population are comparable to those of the UK screening population as a whole, where
the average breast thicknesses were found to be 54.1mm and 56.8mm for the CC and MLO view respectively [189]. These are based on the system-indicated value of compressed breast thickness which typically underestimates the breast thickness at the chest wall [143, 144].

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean (mm)</th>
<th>Standard Deviation (mm)</th>
<th>Median (mm)</th>
<th>Minimum (mm)</th>
<th>Maximum (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>1208</td>
<td>57.3</td>
<td>11.2</td>
<td>57.6</td>
<td>22.8</td>
<td>93.8</td>
</tr>
<tr>
<td>LMLO</td>
<td>1178</td>
<td>60.4</td>
<td>13.1</td>
<td>60.8</td>
<td>21.8</td>
<td>102.3</td>
</tr>
<tr>
<td>RCC</td>
<td>990</td>
<td>57.5</td>
<td>10.8</td>
<td>57.9</td>
<td>20.8</td>
<td>90.0</td>
</tr>
<tr>
<td>RMLO</td>
<td>1172</td>
<td>59.9</td>
<td>12.9</td>
<td>60.2</td>
<td>21.0</td>
<td>99.4</td>
</tr>
</tbody>
</table>

*Table 7.1: Compressed breast thickness measured by the Manchester Method, by view*

The histogram of compressed breast thickness for the LCC view is shown in Figure 7.2. Although there are some differences in the average and spread of data for each view, the frequency distributions for each are very similar and therefore only one representative histogram is included. The greatest number of data points exists for the LCC view, hence this view is used as the example. This is also the case for breast area, volume and density.

![Histogram of compressed breast thickness in the LCC view](image)

*Figure 7.2: Histogram of compressed breast thickness in the LCC view*

Figure 7.2 shows that compressed breast thickness is normally distributed. This is confirmed by the linear normal probability plot and a skewness of -0.08. The mean (57.3mm) and median (57.6mm) values are also very close.
7.3.2.2 Breast area

For the 880 women with data for all 4 views, breast area was calculated for each woman as the mean of all 4 views. Within this sample, the median breast area was 151.7cm² (interquartile range 118.4 – 189.2cm²). The breakdown by view is shown in Table 7.2 and the histogram of breast area for the LCC view is shown in Figure 7.3.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean (cm²)</th>
<th>Standard Deviation (cm²)</th>
<th>Median (cm²)</th>
<th>Minimum (cm²)</th>
<th>Maximum (cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>1208</td>
<td>155.4</td>
<td>66.9</td>
<td>142.8</td>
<td>39.4</td>
<td>484.0</td>
</tr>
<tr>
<td>LMLO</td>
<td>1178</td>
<td>170.5</td>
<td>63.2</td>
<td>160.1</td>
<td>42.6</td>
<td>453.2</td>
</tr>
<tr>
<td>RCC</td>
<td>990</td>
<td>159.8</td>
<td>73.2</td>
<td>145.3</td>
<td>29.8</td>
<td>490.5</td>
</tr>
<tr>
<td>RMLO</td>
<td>1172</td>
<td>169.7</td>
<td>66.4</td>
<td>158.3</td>
<td>52.5</td>
<td>480.2</td>
</tr>
</tbody>
</table>

Table 7.2: Breast area, by view

Figure 7.3 shows that breast area is not normally distributed but is skewed to the left. This is confirmed by the non-linear normal probability plot and a skewness of 1.37. The median value (142.8cm²) is therefore lower than the mean (155.4cm²).
7.3.2.3 Breast volume

For the 880 women with data for all 4 views, breast volume was calculated for each woman as the mean of all 4 views. Within this sample, the median breast volume was 768.5cm³ (interquartile range 554.0 – 1091.4cm³). The breakdown by view is shown in Table 7.3 and the histogram of breast volume for the LCC view is shown in Figure 7.4.

<table>
<thead>
<tr>
<th>View</th>
<th>n</th>
<th>Mean (cm³)</th>
<th>Standard Deviation (cm³)</th>
<th>Median (cm³)</th>
<th>Minimum (cm³)</th>
<th>Maximum (cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>1208</td>
<td>808.4</td>
<td>448.5</td>
<td>710.4</td>
<td>88.1</td>
<td>3127.2</td>
</tr>
<tr>
<td>LMLO</td>
<td>1178</td>
<td>950.4</td>
<td>516.9</td>
<td>845.0</td>
<td>93.6</td>
<td>3398.4</td>
</tr>
<tr>
<td>RCC</td>
<td>990</td>
<td>835.9</td>
<td>482.1</td>
<td>735.1</td>
<td>81.5</td>
<td>3068.7</td>
</tr>
<tr>
<td>RMLO</td>
<td>1172</td>
<td>940.9</td>
<td>532.1</td>
<td>836.9</td>
<td>106.4</td>
<td>3329.5</td>
</tr>
</tbody>
</table>

Table 7.3: Breast volume, by view

Figure 7.4 shows that breast volume is not normally distributed but is skewed to the left. This is confirmed by the non-linear normal probability plot and a skewness of 1.46. The median value (710.4cm³) is therefore lower than the mean (808.4cm³).
7.3.2.4 Breast density

For the 880 women with data for all 4 views, absolute glandular volume and percentage breast density were calculated for each woman as the mean of all 4 views. Within this sample, the median glandular tissue volume was 60.1 cm$^3$ (interquartile range 42.2 – 86.3 cm$^3$). The median percentage breast density was 8.4% (interquartile range 4.9 – 14.2%). The breakdown by view is shown in Table 7.4 and histograms of glandular volume and percentage breast density for the LCC view are shown in Figures 7.5 and 7.6 respectively.

<table>
<thead>
<tr>
<th>Glandular volume (cm$^3$)</th>
<th>n</th>
<th>Mean (cm$^3$)</th>
<th>Standard Deviation (cm$^3$)</th>
<th>Median (cm$^3$)</th>
<th>Minimum (cm$^3$)</th>
<th>Maximum (cm$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>1208</td>
<td>72.7</td>
<td>44.1</td>
<td>62.9</td>
<td>12.6</td>
<td>417.1</td>
</tr>
<tr>
<td>LMLO</td>
<td>1178</td>
<td>72.4</td>
<td>52.6</td>
<td>58.9</td>
<td>7.9</td>
<td>484.3</td>
</tr>
<tr>
<td>RCC</td>
<td>990</td>
<td>70.5</td>
<td>44.8</td>
<td>60.5</td>
<td>9.4</td>
<td>461.3</td>
</tr>
<tr>
<td>RMLO</td>
<td>1172</td>
<td>71.2</td>
<td>46.7</td>
<td>60.3</td>
<td>10.5</td>
<td>395.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage breast density (%)</th>
<th>n</th>
<th>Mean (%)</th>
<th>Standard Deviation (%)</th>
<th>Median (%)</th>
<th>Minimum (%)</th>
<th>Maximum (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>1208</td>
<td>11.8</td>
<td>8.7</td>
<td>9.6</td>
<td>0.8</td>
<td>57.9</td>
</tr>
<tr>
<td>LMLO</td>
<td>1178</td>
<td>10.3</td>
<td>8.8</td>
<td>7.7</td>
<td>0.5</td>
<td>55.8</td>
</tr>
<tr>
<td>RCC</td>
<td>990</td>
<td>11.3</td>
<td>8.7</td>
<td>9.1</td>
<td>0.9</td>
<td>54.9</td>
</tr>
<tr>
<td>RMLO</td>
<td>1172</td>
<td>10.3</td>
<td>8.5</td>
<td>7.8</td>
<td>0.6</td>
<td>51.7</td>
</tr>
</tbody>
</table>

Table 7.4: Glandular volume and percentage breast density, by view

Figures 7.5 and 7.6 show that breast density is not normally distributed but is skewed to the left. This is confirmed by the non-linear normal probability plots and lower median than mean (glandular volume: skewness = 2.41, mean = 72.7 cm$^3$ and median = 62.9 cm$^3$; percentage breast density: skewness = 1.47, mean = 11.8% and median = 9.6%).
Figure 7.5: Histogram of glandular tissue volume in the LCC view

Figure 7.6: Histogram of percentage breast density in the LCC view
7.3.3 Variation with view

Figures 7.7, 7.9, 7.12 and 7.13 show box and whisker plots of compressed breast thickness, breast volume, glandular volume and percentage breast density respectively, for each view. The dark line in the middle of the boxes is the median; the bottom and top of the boxes indicate the 25th and 75th percentiles respectively. The whiskers extend to 1.5 times the height of the box or, if no case has a value in that range, to the minimum or maximum values. The circular points are outliers and the stars are extreme outliers, which are cases that have values more than three times the height of the boxes.

Compressed breast thickness data are distributed normally (Figure 7.2) so approximately 95% of the data lie between the whiskers. Across all 4 views, there are 19 out of 880 points (2.2%) that the statistics software classes as outliers. However, these are distributed uniformly above and below the boxes and all points lie between 20 and 105 mm, which is the range that would be expected.

As discussed in Section 7.3.2.1, there is a difference in compressed breast thickness between the CC and MLO views which is expected due to the inclusion of pectoral muscle in the MLO view. Table 7.5 shows that this difference is statistically significant. Interestingly, the difference in compressed breast thickness between left and right breasts is not statistically significant in the CC view but is significant in the MLO view and consequently the difference between average left and
right compressed breast thickness is also significant. However, the mean difference and t values are low and although the 95% confidence interval does not include zero, the range is small. It is thought that in this case, the p-value is only statistically significant because of the large sample size.

<table>
<thead>
<tr>
<th></th>
<th>Mean Difference</th>
<th>95% Confidence Interval</th>
<th>t</th>
<th>df (n-1)</th>
<th>p-value (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC – LMLO</td>
<td>-3.26</td>
<td>-3.63 -2.88</td>
<td>-16.93</td>
<td>1148</td>
<td>0.000</td>
</tr>
<tr>
<td>RCC – RMLO</td>
<td>-2.94</td>
<td>-3.38 -2.50</td>
<td>-13.10</td>
<td>927</td>
<td>0.000</td>
</tr>
<tr>
<td>LCC – RCC</td>
<td>0.21</td>
<td>-0.08 0.49</td>
<td>1.43</td>
<td>965</td>
<td>0.152</td>
</tr>
<tr>
<td>LMLO – RMLO</td>
<td>0.44</td>
<td>0.21 0.67</td>
<td>3.79</td>
<td>1126</td>
<td>0.000</td>
</tr>
<tr>
<td>Left – Right</td>
<td>0.33</td>
<td>0.11 0.54</td>
<td>2.94</td>
<td>879</td>
<td>0.003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean Difference</th>
<th>95% Confidence Interval</th>
<th>t</th>
<th>df (n-1)</th>
<th>p-value (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC – LMLO</td>
<td>-17.65</td>
<td>-18.90 -16.39</td>
<td>-27.56</td>
<td>1148</td>
<td>0.000</td>
</tr>
<tr>
<td>RCC – RMLO</td>
<td>-17.82</td>
<td>-19.38 -16.26</td>
<td>-22.42</td>
<td>927</td>
<td>0.000</td>
</tr>
<tr>
<td>LCC – RCC</td>
<td>0.35</td>
<td>-1.26 1.95</td>
<td>0.42</td>
<td>965</td>
<td>0.672</td>
</tr>
<tr>
<td>LMLO - RMLO</td>
<td>0.28</td>
<td>-0.95 1.50</td>
<td>0.45</td>
<td>1126</td>
<td>0.655</td>
</tr>
<tr>
<td>Left - Right</td>
<td>0.27</td>
<td>-1.03 1.58</td>
<td>0.41</td>
<td>879</td>
<td>0.680</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean Difference</th>
<th>95% Confidence Interval</th>
<th>t</th>
<th>df (n-1)</th>
<th>p-value (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC – LMLO</td>
<td>-158.25</td>
<td>-167.97 -148.52</td>
<td>-31.93</td>
<td>1146</td>
<td>0.000</td>
</tr>
<tr>
<td>RCC – RMLO</td>
<td>-159.75</td>
<td>-172.08 -147.43</td>
<td>-25.43</td>
<td>927</td>
<td>0.000</td>
</tr>
<tr>
<td>LCC – RCC</td>
<td>3.45</td>
<td>-7.29 14.18</td>
<td>0.63</td>
<td>964</td>
<td>0.529</td>
</tr>
<tr>
<td>LMLO - RMLO</td>
<td>5.47</td>
<td>-3.38 14.31</td>
<td>1.21</td>
<td>1125</td>
<td>0.226</td>
</tr>
<tr>
<td>Left - Right</td>
<td>4.25</td>
<td>-4.88 13.37</td>
<td>0.91</td>
<td>879</td>
<td>0.361</td>
</tr>
</tbody>
</table>

*Table 7.5: Variation in compressed breast thickness, area and volume, by view*

For breast area, the difference between left and right breasts is not statistically significant. However, the difference between MLO and CC views is significant. This result was surprising, especially as the pectoral muscle is removed by the operator and does not contribute to the breast area. However, the illustrative example in Figure 7.8 suggests that this is a genuine result, based
on differences in radiographic positioning [190]. Figure 7.8a shows schematic drawings of a compressed breast positioned in the CC view and 7.8b shows the same compressed breast in the MLO view. The areas of the CC view and the pectoral muscle have been roughly superimposed on the MLO view in Figure 7.8b and it can be seen that there is still a small amount of additional breast tissue, compared to the CC view.

Breast volume is dependent on thickness and area so inevitably, the breast volume is also significantly greater in the MLO view than the CC view but there is no significant difference between the left and right breasts, as shown in Table 7.5 and Figure 7.9.
Breast volume data are not distributed normally but skewed to the left (Figure 7.4). The whiskers in Figure 7.9 therefore extend to the minimum value and to 1.5 times the height of the box in the upper direction. The statistics software defines 49 out of 880 points (5.6%) as outliers and even 20 points (2.3%) as extreme outliers. For this reason, the association between breast volume and compressed breast thickness was examined. Using non-linear regression analysis, the relationship is found to be approximately quadratic ($y = 0.286x^2 - 2.344x$, $R^2 = 0.923$), as shown in Figure 7.10. Only the LMLO view is shown; this is considered representative of all views.

![Figure 7.10: Relationship between breast volume and compressed breast thickness in the LMLO view](image)

There is a strong and significant correlation between compressed breast thickness and breast volume. The Spearman correlation coefficient (Spearman’s rho, $\rho$) is used because of the non-normal distribution of breast volume data. For the LCC view, $\rho = 0.695$ ($p < 0.01$) and for the LMLO view, $\rho = 0.846$ ($p < 0.01$). The fact that the highest values of breast volume are associated with very high compressed breast thickness, suggests that they are true data points and the maximum breast volume of 3,398 cm$^3$ is considered to be a genuine result. The corresponding mammogram is shown in Figure 7.11. The field size is 24x30cm and the maximum compressed breast thickness is 91.9mm.
Figure 7.11: LMLO mammogram corresponding to the largest breast volume (3,398 cm³)

Table 7.6 shows that interestingly, glandular tissue volume remains approximately constant, regardless of view, with no significant differences observed between CC and MLO or between left and right breasts.

<table>
<thead>
<tr>
<th>Glandular tissue volume (cm³)</th>
<th>Mean Difference</th>
<th>95% Confidence Interval</th>
<th>t</th>
<th>df (n-1)</th>
<th>p-value (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC – LMLO</td>
<td>-1.09</td>
<td>-3.11 – 0.92</td>
<td>-1.07</td>
<td>1144</td>
<td>0.287</td>
</tr>
<tr>
<td>RCC – RMLO</td>
<td>-1.60</td>
<td>-3.36 – 0.16</td>
<td>-1.78</td>
<td>925</td>
<td>0.075</td>
</tr>
<tr>
<td>LCC – RCC</td>
<td>1.13</td>
<td>-0.18 – 2.43</td>
<td>1.69</td>
<td>962</td>
<td>0.091</td>
</tr>
<tr>
<td>LMLO – RMLO</td>
<td>1.48</td>
<td>-0.27 – 3.23</td>
<td>1.66</td>
<td>1124</td>
<td>0.097</td>
</tr>
<tr>
<td>Left – Right</td>
<td>1.09</td>
<td>-0.25 – 2.42</td>
<td>1.60</td>
<td>878</td>
<td>0.110</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage breast density (%)</th>
<th>Mean Difference</th>
<th>95% Confidence Interval</th>
<th>t</th>
<th>df (n-1)</th>
<th>p-value (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC – LMLO</td>
<td>1.70</td>
<td>1.49 – 1.92</td>
<td>15.89</td>
<td>1144</td>
<td>0.000</td>
</tr>
<tr>
<td>RCC – RMLO</td>
<td>1.40</td>
<td>1.18 – 1.63</td>
<td>12.18</td>
<td>925</td>
<td>0.000</td>
</tr>
<tr>
<td>LCC – RCC</td>
<td>0.02</td>
<td>-0.18 – 0.21</td>
<td>0.17</td>
<td>962</td>
<td>0.863</td>
</tr>
<tr>
<td>LMLO – RMLO</td>
<td>-0.02</td>
<td>-0.22 – 0.18</td>
<td>-0.21</td>
<td>1124</td>
<td>0.832</td>
</tr>
<tr>
<td>Left – Right</td>
<td>-0.02</td>
<td>-0.18 – 0.14</td>
<td>-0.27</td>
<td>878</td>
<td>0.788</td>
</tr>
</tbody>
</table>

Table 7.6: Variation in glandular tissue volume and percentage breast density, by view
The difference in percentage breast density between left and right breasts is not statistically significant. However, the difference in breast density between CC and MLO views is statistically significant for both breasts, with breast density being higher in the CC view. The same observations have been made for radiologist-assessed breast density by area [116, 191]. Given that glandular volume is not significantly different between views, it seems likely that the significant difference in percentage breast density between CC and MLO views stems from significant differences in measured breast volume, which are themselves the result of differences in compressed breast thickness and area.

Glandular tissue volume and percentage breast density data are not distributed normally but skewed to the left (Figures 7.5 and 7.6). The whiskers in Figures 7.12 and 7.13 therefore extend to the minimum value and to 1.5 times the height of the box in the upper direction. There are several points that the statistics software defines as outliers (4.4% and 6.6% of points for glandular volume and percentage breast density respectively), and even extreme outliers (4.9% and 2.3% respectively). It is possible that these could be genuine anomalies in the data due to software failure. An investigation into the extreme outliers is presented below.

![Figure 7.12: Box and whisker plots of glandular tissue volume by view](image-url)
There were no women with breast implants in the sample. For the 10 women with the greatest glandular tissue volume (> 250cm³), all but one also had very high percentage breast density (> 28%, 94th percentile). One woman had percentage breast density of 11.6% and in this case the high glandular volume could be explained by a very high breast volume (2,214cm³, 98th percentile). For the 10 women with the greatest percentage density (> 44%), 6 were also found to exhibit very high glandular volume (> 150cm³, 95th percentile). For the remaining four women, the high breast density but lower glandular volume could simply be explained by low breast volume (< 168.4cm³, 2nd percentile).

Figure 7.14 illustrates a case where the percentage breast density was classed as extremely high (exceeding 99th percentile) but glandular volume was not (below 86th percentile). Figures 7.15 and 7.16 are examples of high breast density (exceeding 96th percentile) and high glandular volume (exceeding 97th percentile). Details are shown in Table 7.7 below. Figure 7.15 is highlighted as this has the highest percentage breast density in the study (mean of all 4 views = 55.1%).

<table>
<thead>
<tr>
<th>Figure</th>
<th>View</th>
<th>Breast thickness (mm)</th>
<th>Breast volume (cm³)</th>
<th>Glandular volume (cm³)</th>
<th>Percentage breast density (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.14</td>
<td>RCC</td>
<td>38.9</td>
<td>226.8</td>
<td>106.0</td>
<td>46.7</td>
</tr>
<tr>
<td>7.15</td>
<td>LCC</td>
<td>33.0</td>
<td>323.6</td>
<td>187.5</td>
<td>57.9</td>
</tr>
<tr>
<td>7.16</td>
<td>RMLO</td>
<td>86.9</td>
<td>1364.0</td>
<td>395.1</td>
<td>29.0</td>
</tr>
</tbody>
</table>

Table 7.7: Breast thickness, volume and density for the cases shown in Figures 7.14 – 7.16
Referring to the original mammogram images in Figures 7.14 and 7.15, it can be seen that a large proportion of the total breast area is occupied by bright (radiodense) tissue and visually, these breasts would be classified as having high mammographic density. The glandular thickness plots are a good reflection of the original mammograms and the stepwedge intensity plots contain the full range of grey levels, both of which are good indications that the method has been successful and the high breast density results are genuine. However, they both have low compressed breast thickness (below the 6th and 2nd percentile for 7.14 and 7.15 respectively). It is therefore likely that
the x-ray exposure would have used relatively low kV and mAs and this is supported by the appearance of the stepwedge on the original mammogram images. The lowest steps are clearly discernible and in both examples, the intensity plot shows that the minimum pixel intensity (grey level) is reached by Step 24.

Figure 7.15: Top: original mammogram and glandular thickness map for a breast with 57.9% density and glandular volume of 187.5cm$^3$; Bottom: stepwedge pixel intensity plot
Figure 7.16 is an example of the method being pushed beyond its acceptable limits. The compressed breast thickness of 87mm exceeds the 99th percentile and is much greater than the previous two examples. It is therefore likely that the x-ray exposure would have used high kV and mAs. This is reflected in the appearance of the step wedge on the original mammogram image where it is very hard to resolve the greylevels of the lower steps. Furthermore, the step wedge intensity plot does not follow the expected curve and shows that the full range of grey levels has
not been covered, reaching a minimum of 1600 (equivalent to an optical density of 1.6) at the maximum step height of 14mm. The glandular thickness map is a reasonable reflection of the original mammogram but the central portion is saturated. Because the stepwedge does not cover the full range of grey levels in the mammogram, there is only limited confidence that this result is genuine. The percentage breast density calculated by the software may in fact be too low, despite the fact that visually, the mammographic density of this breast is lower than Figures 7.14 and 7.15.

7.3.4 Variation of breast density with compressed breast thickness and breast volume

Figure 7.17 shows the relationship between percentage breast density with compressed breast thickness (left) and breast volume (right) for the LCC view. Spearman's rho is used to examine the correlation between variables because of the non-normal distribution of breast density data. As expected, breast density decreases significantly with compressed breast thickness, $\rho = -0.636$ ($p < 0.01$) and breast volume, $\rho = -0.727$ ($p < 0.01$).

It is interesting to observe that the absolute volume of glandular tissue remains approximately constant with compressed breast thickness and volume, as shown in Figure 7.18. If anything, the spread of data decreases with increasing breast volume. There are very weak negative correlations for compressed breast thickness, $\rho = -0.205$ ($p < 0.01$) and breast volume, $\rho = -0.080$ ($p < 0.01$). The statistical significance is thought to be due to the large number of data points.
7.3.5 Method failures

1,401 women completed the questionnaire and consent form. 10 film packets were returned to Bolton urgently as the women had been recalled for assessment; these were not subsequently retrieved. The films of 1,391 women were therefore available for digitisation. There were 3 women with implants whose films were not digitised and 8 women who had had a mastectomy who therefore only had 2 films. In total, there were 5,536 films.

Of these, 408 films (from 102 women) were not digitised as it would have been impossible to run the software on the images. The reasons were a missing stepwedge because the breast was too big (41 cases), the stepwedge being misplaced and being obscured by the ID label (60 cases) or a year marker being placed over the stepwedge (1 case). These problems occurred on all 4 views.

Within a number of film packets, there were individual images that were not digitised, or the software was not successful. Out of 5,536 films, the number of missing images for each view was:

<table>
<thead>
<tr>
<th>View</th>
<th>LCC</th>
<th>LMLO</th>
<th>RCC</th>
<th>RMLO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of missing images</td>
<td>185</td>
<td>215</td>
<td>402</td>
<td>221</td>
</tr>
</tbody>
</table>

Table 7.8: Number of missing images for each view

The RCC view is highlighted because there are many more images missing in this view. The reason for this is that the radiographers are required to place an address label on this film. Unfortunately this was often placed over the stepwedge. The label could not easily be removed and would leave a residue on the film which would damage the digitiser. Obviously the sticker could not be placed over the breast and on some occasions, the breast area was so great that the sticker could only be placed over the stepwedge. However, these occasions were rare. It is
thought that because appointment times are so short and the use of the stepwedge method is not routine, the radiographers did not have time to think about where to place the sticker. This is supported by the fact that this problem was much worse on the Bolton van, where the time pressure is greater than that at a static site, simply because of increased numbers attending for screening, and the lack of space in the waiting area. Of the 223 RCC images with a sticker over the stepwedge, 140 of these were from the van at Bolton. Given that only 325 women attended at this site and assuming that they all had a RCC view taken, this means that 43% of the RCC films from this site could not be analysed. Although this caused the method to fail, it is an issue that could be addressed and eliminated in future studies.

In total, 10 reasons of method failure were identified which are listed in Figure 7.19. The pie-charts in Figure 7.19 show the proportion of cases failing due to each reason. Two pie-charts are shown for the RCC view; the top one includes the reason “sticker over stepwedge.” The lower chart shows how this would change if this reason was removed, given that this problem was exclusive to the RCC view (with the exception of a single RMLO view). This pie-chart for the RCC view is shown alongside the pie-chart for the RMLO view. It can be seen that the distribution of method failures is very similar; this is also true for the LCC and LMLO views.

Although some failures are anticipated, it is disappointing if these exceed expectations. The purpose of this feasibility study was to examine causes of failure and fortunately many of these could be prevented with better education of the staff involved. It is hoped that fewer failures would be observed if the method became routine. An example of each type of failure and a discussion of prevention strategies is presented on pages 169 - 171. Note that there are no examples of the sticker over the stepwedge as none of these films were digitised. It is important to note that there was only one film where the failure of the method also meant that diagnosis on the breast image was impeded. All four views from this examination are shown in Figure 7.20; fortunately only one view is affected. On the LMLO view (thickness = 68mm, area = 440cm², volume = 2,716cm³) the stepwedge obscured the breast which meant that it had been placed on top of the compression paddle.
Figure 7.19: Pie-charts showing the proportion of cases for each method failure. The top chart for the RCC view includes the reason “sticker over stepwedge”. If this reason is removed, the distribution of method failures is very similar for each view.

Figure 7.20: On the LMLO image (shown by red asterisk), the stepwedge obscures the breast. This is the only view affected within this examination and within the whole sample.
Reasons for failure:

1) **Sticker over stepwedge**

Better education of staff involved in study. Even if the breast area was too great to find room for the address label, it could have been stuck to the back of the film or to one of the other views. *This image shows the sticker in the top left-hand corner. Fortunately it does not obscure the useful area of the stepwedge but it just encroaches on the breast. Analysis of breast density was still possible.*

2) **Stepwedge in ID label**

Better education of staff involved in study. If the breast area was particularly great (as in the image shown), it may not even be possible to fit the stepwedge alongside the breast. It would either get pushed out of the way, thereby going into the ID label, or would have to be flipped out of the field of view. Therefore, not all failures could be prevented.

3) **No stepwedge (breast too big)**

Very few of these failures are preventable. In some cases, it may have been possible to use a 24x30cm film instead of 18x24cm. However, changing the bucky increases the length of the appointment time.

4) **Artefact on image**

Artefacts can occur on clinical images so are not strictly a failure of the method. Fortunately, they are very rare. In this case, the ‘artefact’ is a pacemaker, which makes it impossible to assess breast density.
Reason for failure:

5) Marker sheet missing

Better education of staff involved in study to ensure that the sheet is always in place.

However, in a small number of these cases, the compressed breast thickness may have been so great that the markers will be off the edge of the film, simply due to magnification geometry.

6) Stepwedge obscures breast

The breast is so large (this is a 24x30cm film) that the stepwedge has been placed on top of the compression paddle. Fortunately, this only occurred on one image.

7) Software failure: breast edge detection

This tended to occur when the breast was large and the breast edge almost met the stepwedge and / or other markers (top image).

On the bottom image, the breast has been poorly positioned and there is abdominal fat included at the chest wall edge which confused the edge detection algorithm.

It would be possible to repeat the analysis on these films using manual breast edge detection, but this could be time-consuming in a large scale study.
Reason for failure:

8) No markers on one edge

Very few of these failures could be prevented. If the breast is above a certain thickness (9.4cm on 18x24cm films and 10.4cm on 24x30cm films), magnification geometry means that the markers will be off the edge of the film, on one or both edges.

9) No stepwedge (breast too small)

These failures are unpreventable. If the compressed breast thickness is less than 14mm, the stepwedge cannot be included in the field of view. This only affected one image within the sample.

10) Year marker obscures stepwedge

Better education of staff involved in the study.

7.3.6 Comments from invited women

A log of twenty one telephone calls was recorded in response to the study documentation. The queries and responses are shown in Table 7.9.
<table>
<thead>
<tr>
<th>Query (Q) and Answer (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q. 6 women wished to change their appointment time or attend a different screening centre.</td>
</tr>
<tr>
<td>A. Their call was redirected to Bolton Breast Screening Centre.</td>
</tr>
<tr>
<td>Q. A number of women wished to check their eligibility for taking part in the study due to:</td>
</tr>
<tr>
<td>• previous breast cancer (1)</td>
</tr>
<tr>
<td>• mastectomy (2)</td>
</tr>
<tr>
<td>• implants (1)</td>
</tr>
<tr>
<td>• aged over 70 (1)</td>
</tr>
<tr>
<td>A. Study eligibility was confirmed for all women except the one with implants. However, she was still encouraged to attend for routine breast screening. Women with previous cancer and mastectomy were asked to note this on their questionnaire.</td>
</tr>
<tr>
<td>Q. 1 woman sought clarification on the implications of taking part in the study.</td>
</tr>
<tr>
<td>A. The study details and implications were discussed in detail.</td>
</tr>
<tr>
<td>Q. 1 woman requested to take part as her mother had received an invitation. She had been referred for mammograms since the age of 34 as she was considered high-risk based on a history of breast lumps. Her last mammogram was 3 months ago.</td>
</tr>
<tr>
<td>A. The woman was thanked for her interest in the study but informed that she was too young and her last mammogram had been too recent to justify taking part. Some literature on breast density was sent to her home address.</td>
</tr>
<tr>
<td>Q. 1 woman was concerned by additional radiation dose.</td>
</tr>
<tr>
<td>A. She was reassured that the study required no additional radiation dose as the information could be gained from the standard mammographic views.</td>
</tr>
<tr>
<td>Q. 3 women requested the results of study.</td>
</tr>
<tr>
<td>A. The women were thanked for their interest and informed that results would be provided but analysis would take some time and results would not be personalised. Some literature on breast density was provided in the meantime.</td>
</tr>
<tr>
<td>Q. 2 women received a study invitation but not a screening invitation.</td>
</tr>
<tr>
<td>A. An apology was given and it was explained that this was an administrative error. The women were asked to contact Bolton Breast Screening Centre if they thought they were due for a mammogram.</td>
</tr>
<tr>
<td>Q. 1 woman was concerned that she had received the invitation because of an abnormality on a previous mammogram.</td>
</tr>
<tr>
<td>A. She was reassured that the only criterion for participant selection was that they were due to attend for routine breast screening in the next 6 months.</td>
</tr>
<tr>
<td>Q. 1 woman was happy to take part but was offended by the ethnicity question as “White British” was one of the last options.</td>
</tr>
<tr>
<td>A. An attempt was made to reassure the woman that the ethnicity listing was alphabetical and reflected the order and categories of the Census. An apology was given for offence caused.</td>
</tr>
</tbody>
</table>

Table 7.9: Queries regarding the breast density study and responses given
7.4 Discussion

One of the main aims of the study was to assess whether the current implementation of the Manchester Method for breast density measurement is practical for use in the screening programme, where the average appointment time is 6 minutes or less. The method had previously only been assessed on a small sample of women taking part in a lifestyle study. This earlier study used a PTFE stepwedge which was far larger and more difficult to attach to the breast support platform, making it unsuitable for use in the screening programme. The compact aluminium stepwedge can remain clipped to the breast support platform for almost all examinations. There is no need to reposition it between views and it can be easily flipped out of the field of view on the rare occasions where it will not fit alongside the breast.

Feedback from the radiographers involved in the study suggests that the method did not negatively impact on the screening appointment time. However, there were a number of method failures which could be attributed to radiographer error, for example, placing an address label over the stepwedge, not including the marker sheet on the compression paddle, and misaligning the stepwedge so that it was obscured by the ID label. If the radiographers were under less time pressure, it is possible that there would be fewer failures for these reasons. This is supported by the observation that a higher percentage of failures occurred on the van rather than at the static site. Therefore, although the stepwedge method can practically be implemented in the screening programme, without detrimental effect on clinical workflow, there is certainly room for improvement in terms of the success rate of the method. It is thought that better staff education and possibly supervision at the start of the study would eliminate these failures in future studies.

In total, data analysis was not possible on 18% of films. This result is skewed by the large number of RCC views that could not be used due to an address label being placed over the stepwedge. As discussed above, a large proportion of method failures would be preventable in future studies by better radiographer education. Images where the breast edge detection algorithm failed could be marked-up manually. The only method failures that cannot be addressed are those with no room for the stepwedge due to the breast area being too large (3.9%) or compressed breast thickness too small (0.1%) and those where markers are missing on at least one edge (1.1%). Some films where the marker sheet was missing or the stepwedge went into the ID label may have also been unpreventable due to large compressed breast thickness or area. Taking these into account, it is thought that the method will genuinely fail on approximately 6% of images. However, this does not mean that the method would fail on 6% of women. Given that the difference in glandular volume and breast density between left and right breasts is not statistically significant (Table 7.6), a reliable measure of breast density could be obtained using the data from one breast only. It is likely that there would have been even more unavoidable failures related to breast size if the previous PTFE stepwedge was used.
Another aim of the study was to find out if the method could adequately cope with the ranges in compressed breast thickness and density expected within the screening population. The results presented in Section 7.3.2 certainly indicate that breast density measurements can be made over the full range of compressed breast thicknesses anticipated (although the magnification markers will only be visible on the image up to a compressed breast thickness of 94mm and 104mm for 18x24cm and 24x30cm field sizes respectively). However, an analysis of outliers revealed a lack of confidence in the accuracy of density measurements in breasts that are both extremely thick and highly dense, such as the example in Figure 7.16. In this case, the grey levels within the stepwedge did not cover the full range of grey levels expected within correctly exposed breast tissue. The RMLO view was presented in Figure 7.16; the LCC view for the same woman is highlighted in Figure 7.21 below, which shows the relationship between breast density and thickness. It can be seen that this is the most extreme outlier in the sample. The number of cases that will be affected by this issue is likely to be very low and is estimated as follows. Based on an evaluation of the optical density profiles of the stepwedge images collected for calibration, it would seem reasonable to assume that there is a threshold of approximately 40% density for 60mm breasts and 30% density for 70mm breasts, beyond which the stepwedge fails to cover the range of optical densities expected clinically. A linear extrapolation of these points is shown as the blue line on Figure 7.21. It can be seen that this excludes only 4 cases, with another 7 lying on, or very close to this line. Given that there are 1,208 data points for the LCC view, this represents a maximum of 0.9% of cases where the breast density measurement is potentially incorrect.

Figure 7.21: Percentage breast density versus compressed breast thickness; the blue line is an estimation of the point at which the accuracy of breast density measurements becomes unreliable
The median percentage breast density within the sample was 8.4% (interquartile range 4.9 – 14.2%) and the median glandular volume was 60.1cm³ (interquartile range 42.2 – 86.3cm³). There was no significant difference in glandular volume or percentage breast density between the left and right breasts. Additionally, the correlation between the left and right breasts (using the Pearson correlation coefficient, r) was found to be significant (p < 0.001) and stronger than that quoted for other volumetric techniques [140, 153, 157], as shown in Table 7.10.

<table>
<thead>
<tr>
<th>Percentage breast density (%)</th>
<th>Manchester Method</th>
<th>SXA [157]</th>
<th>SMF [140]</th>
<th>Quantra™ [153]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glandular volume (cm³)</td>
<td>0.90</td>
<td>0.77</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 7.10: Pearson correlation coefficients between left and right breasts
SMF: Standard Mammographic Form; SXA: Single X-ray Absorptiometry

However, the difference in percentage breast density between the CC and the MLO view was significant, despite being closely correlated (r = 0.92, p < 0.001). On average, the CC view was found to be denser than the MLO view, an observation also made using radiologist visual-assessment [116, 191]. This suggests that it may not be adequate to measure density for one view only and for this reason, correlation between breast density and risk-related factors will be examined using LCC and LMLO data. There is evidence to suggest that the strength of the association between breast density and risk increases when considering both the CC and MLO views compared to the MLO view only [116]. However, this paper was based on a study using radiologist visual assessment of density and other studies using automated techniques have shown that it is possible to use one view only [22].

Interestingly, glandular volume was not significantly different between mammographic views and it therefore seems probable that the significant difference in percentage breast density is the result of differences in breast volume. Breast volume is itself dependent upon compressed breast thickness and area, which were also significantly different for the CC and MLO views. It is acknowledged that the MLO view contains the axillary tail and may also include more tissue at the chest wall edge of the breast; these areas are thought to be primarily fat [130].

The consent rate of 24% was disappointing but appears to be in line with similar studies in the literature [192]. Despite the low consent rate, the sample size was great enough to examine breast density distribution and correlate breast density with a number of risk factors with a reasonable degree of statistical confidence. The anticipated rate was 50% and radiographers questioned after the first few weeks of the study felt that this was being achieved. This may have been slightly ambitious, given that the requirement in the NHSBSP is ≥ 70% of invited women to attend for screening, with a target of 80% [41]. In Bolton, the overall uptake rate from April 2007.
to March 2008 was 74% (51% at the prevalent round, 83% at the incident round) [193]. In Australia, the overall uptake rate is only 54.3% (1997 – 1998) [194].

Participation rates in epidemiological studies are declining [195] and in a sample of 355 articles from high impact-factor medical and epidemiology journals in 2003, participation rates as low as 15% were recorded. In over 50% of these articles, no details of participation rates were even provided. Reasons for not participating in a clinical study include difficulty in understanding the study documentation, the purpose of the study being unclear, a lack of perceived value of the study (especially if there is no direct personal benefit) and procrastination or simply forgetting to take part [192]. In the case of the Bolton study, all of these reasons are feasible. The patient information sheet stated that there would be no direct benefit to individuals, although taking part would benefit women in the future, especially those thought to be at a higher risk of developing breast cancer. Participants were also informed that the markers used to measure breast thickness could overly tissue on women with larger breasts. Previous work has shown that the markers only overly the breast tissue in 2.5% of cases and do not overly the same area of tissue on both views [182]. However, a small number of women provided this as the reason for not wishing to take part.

The main reason for not participating is thought to be forgotten questionnaires. This is supported by the fact that the consent rate at the static site was almost three times that on the screening van. This can most likely be attributed to the fact that there was a large reception area at the static site where several women could wait for their appointment and complete an additional questionnaire if they had forgotten to bring theirs. There was sufficient space in the reception office to store a box of spare questionnaires but this was not the case on the van. In addition, there was a receptionist available at all times to answer any questions or concerns they had about the study. Personal contact is thought to be an incentive to taking part in a study [192]. The van was staffed by only two radiographers who move between the reception area, the changing rooms and the examination room.

In future studies women could be provided with a pre-paid envelope for returning their questionnaire. In 2009, this approach was trialled on the same study invitees. A new questionnaire was constructed which contained additional questions on deodorant use, smoking and natural HRT remedies. The questions on menopausal status, age at first child and family history were expanded as these had led to ambiguous responses. Despite being twice as long as the first questionnaire, the response rate was 42%.

The advantages of the Manchester Method are that it is an objective technique and provides a measure of the volume of glandular tissue in addition to volumetric percentage density. These measures should provide more meaningful correlations with factors such as weight and BMI than a subjective assessment of breast density based on area. Correlations with risk-related factors are examined in the next chapter.
8. Relationship between Volumetric Breast Density and Breast Cancer Risk Factors

This chapter examines the relationship between volumetric breast density and those breast cancer risk factors that were collected by questionnaire as part of the study described in Chapter 7. Although much has been published on the relationship between mammographic (area-based) density and risk factors, little information of this nature exists for volumetric breast density.

Collection of risk factor information, combined with a measure of breast density can be used to develop individualised risk prediction models. These models are constantly evolving. Barlow et al [27] recently developed a model including breast density (using the BI-RADS classification [28]) and the use of hormone therapy as additional inputs. Breast density was found to be a statistically significant independent risk factor for breast cancer diagnosis in pre- and post-menopausal women and it is thought that its inclusion in risk prediction models may offer improved accuracy in the identification of women at high risk of developing breast cancer.

An improved risk model could facilitate risk-based screening whereby the frequency of screening is determined by the level of risk. This is beyond the scope of this PhD but is currently being investigated in the PROCAS study [112, 113] (Predicting Risk Of Cancer At Screening) described in Chapter 10 (Future Work). Experiences from the feasibility study at Bolton have been invaluable in informing the design of the PROCAS study, particularly in terms of the questionnaire development and strategies for optimising participation rate.

8.1 Study Questionnaires

The original questionnaire (2007), which will be referred to as Questionnaire 1 gathered information on date of birth, height, weight, age at birth of first child, ages of menarche and menopause, ethnicity and family history of breast cancer (mother or sister only) including the age at which breast cancer was diagnosed in this relative. Information on the use of hormone replacement therapy was taken from the patient’s notes. A Likert scale [188] was associated with each question (‘don't know’, ‘not sure’, ‘quite sure’, ‘certain’) to ascertain the level of confidence in the response. In order to determine the accuracy of self-reported weight, it was anticipated that a sample of the women attending at the static site would be weighed using scales calibrated on a monthly basis and the actual weight recorded alongside the reported value.

In 2009, the Bolton Local Research Ethics Committee approved a substantial amendment to the original study design, which allowed an additional questionnaire to be sent to the invitees of the 2007 feasibility study. There were a number of reasons for doing this. As discussed in Chapter 7, the consent rate for the study (24% overall) was lower than anticipated and it was thought that this
was primarily due to women forgetting to take their questionnaire to their appointment. To investigate this hypothesis, a pre-paid envelope was provided for the second questionnaire and the response rate increased to 42%.

Furthermore, it became apparent during data entry that the responses to some of the questions were open to interpretation. For example, a blank entry for ‘age at menopause’ or ‘age at first child’ could mean that the woman has not started the menopause or has not had children, rather than that she has ignored the question. There was not space on the questionnaire to clarify this. For this reason, these sections were expanded. The second questionnaire will be referred to as Questionnaire 2. Both questionnaires are included in the Appendix.

It was also thought that a four-point confidence scale was unnecessary, especially as ‘not sure’ and ‘don’t know’ have similar meanings. The term ‘quite sure’ was replaced by ‘fairly sure’ as it was felt that this was less ambiguous.

The grant from Genesis is gratefully acknowledged for this research. A condition of receiving this grant was that the questionnaire would also gather information on the use of natural HRT, smoking, deodorant use and postcode. Postcode data would enable the investigation of breast density variations within urban and rural populations and provide an indication of socioeconomic status.

The second questionnaire has therefore facilitated the examination of the relationships between breast density and additional risk factors. It has also permitted analysis of the reliability and consistency of self-reported data.

8.2 Relationship between breast density and breast cancer risk factors

All statistical analyses were carried out using IBM SPSS Version 19.

The correlation of glandular volume and percentage breast density with the continuous risk variables is shown below in Tables 8.1 and 8.2 respectively using data for the LCC, LMLO and left breast (average of LCC and LMLO views). The Spearman rank correlation coefficient (Spearman’s rho, ρ) is used because of the non-parametric distribution of the breast density variables.
### Relationship with glandular volume (cm³), by view

<table>
<thead>
<tr>
<th></th>
<th>LCC Spearman ρ</th>
<th>LMLO Spearman ρ</th>
<th>Left breast Spearman ρ</th>
<th>Significance level (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.139</td>
<td>-0.212</td>
<td>-0.193</td>
<td>0.000</td>
</tr>
<tr>
<td>Weight</td>
<td>-0.136</td>
<td>-0.140</td>
<td>-0.145</td>
<td>0.000</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.165</td>
<td>-0.197</td>
<td>-0.189</td>
<td>0.000</td>
</tr>
<tr>
<td>Age at menarche</td>
<td>-0.032</td>
<td>-0.003</td>
<td>-0.021</td>
<td>0.484</td>
</tr>
<tr>
<td>Age at menopause</td>
<td>0.091</td>
<td>0.105</td>
<td>0.110</td>
<td>0.001</td>
</tr>
<tr>
<td>Years of menstruation</td>
<td>0.098</td>
<td>0.101</td>
<td>0.112</td>
<td>0.001</td>
</tr>
<tr>
<td>Age at first child</td>
<td>0.015</td>
<td>0.089</td>
<td>0.060</td>
<td>0.057</td>
</tr>
</tbody>
</table>

**Table 8.1: Correlation of glandular volume with breast cancer risk factors**

Glandular volume in the left breast shows a significant correlation with all variables (p < 0.001, 2-tailed) except for age at first child and age at menarche. However, the strength of correlation is weak for some variables and statistical significance may only have been reached because of the large number of data points.

### Relationship with percentage breast density (%), by view

<table>
<thead>
<tr>
<th></th>
<th>LCC Spearman ρ</th>
<th>LMLO Spearman ρ</th>
<th>Left breast Spearman ρ</th>
<th>Significance level (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.101</td>
<td>-0.167</td>
<td>-0.138</td>
<td>0.000</td>
</tr>
<tr>
<td>Weight</td>
<td>-0.551</td>
<td>-0.510</td>
<td>-0.535</td>
<td>0.000</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.598</td>
<td>-0.576</td>
<td>-0.594</td>
<td>0.000</td>
</tr>
<tr>
<td>Age at menarche</td>
<td>0.082</td>
<td>0.082</td>
<td>0.075</td>
<td>0.013</td>
</tr>
<tr>
<td>Age at menopause</td>
<td>0.043</td>
<td>0.070</td>
<td>0.068</td>
<td>0.035</td>
</tr>
<tr>
<td>Years of menstruation</td>
<td>0.013</td>
<td>0.037</td>
<td>0.038</td>
<td>0.240</td>
</tr>
<tr>
<td>Age at first child</td>
<td>0.019</td>
<td>0.073</td>
<td>0.051</td>
<td>0.102</td>
</tr>
</tbody>
</table>

**Table 8.2: Correlation of percentage breast density with breast cancer risk factors**

Left breast density shows a significant correlation (p < 0.001, 2-tailed) with age, weight and BMI. A weaker correlation is demonstrated for age at menarche and menopause and the significance level is also lower (p < 0.05, 2-tailed). There is no significant correlation with age at first child or years of menstruation.
In order to determine whether the variables highlighted in the tables above are genuinely correlated with volumetric breast density, or whether the relationship is confounded by other variables, multiple linear regression analysis was carried out. A model of the form

\[ y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \]

was fitted to the data, where \( x_1 \) to \( x_4 \) were age, weight, BMI, and age at menopause, respectively.

In the model for glandular tissue volume, \( x_5 \) was years of menstruation. Multiple linear regression revealed that only age (\( p < 0.0001 \)), weight (\( p < 0.0005 \)) and BMI (\( p < 0.0001 \)) retained significance.

In the multiple linear regression model for percentage breast density, \( x_5 \) was age at menarche. Only age, weight and BMI (all \( p < 0.0001 \)) retained significance.

More detailed analysis is presented in the following sections, using data from the left breast only (taken as the average of the CC and MLO views). In Chapter 7, it was shown that the difference in breast density between the left and right breasts was not significant and there is a greater quantity of data available for the left breast.

### 8.2.1 Age at Examination

Age at examination was recorded for 1,261 participants. The mean age of the sample was 60.0 (standard deviation, s.d. 5.7; range 48 to 78) which is exactly as expected, given that women are invited for breast screening between the ages of 50 – 70, and can self-refer from the age of 70.

Figures 8.1 and 8.2 show the relationship between glandular volume and percentage breast density respectively with age at examination. Because of the non-parametric distribution of breast density variables, a logarithmic scale has been used. Breast density decreases with age, with glandular volume exhibiting a stronger correlation (\( \rho = -0.193 \)) than percentage breast density (\( \rho = -0.138 \)). Although the correlations are significant (\( p < 0.001 \)), the relationships are fairly weak. A possible reason for this is that the majority of the women in the sample are post-menopausal. This was investigated further.
Table 8.3 shows the correlation of glandular volume and percentage breast density with age for pre- and post-menopausal women using data from Questionnaire 1. Note that the question asked only for “age at menopause”. Some women clearly stated that they were pre-menopausal and were placed in this category; a blank response to this question was not assumed to mean that the
A woman was pre-menopausal. This ambiguity was resolved in Questionnaire 2 where menopause status was additionally requested. Unfortunately, the sample sizes were too small to permit meaningful analysis (n = 353, n = 8 and n = 63 for post-, pre- and peri-menopausal women respectively).

<table>
<thead>
<tr>
<th>Sample size, n</th>
<th>Glandular volume (cm$^3$)</th>
<th>Breast density (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spearman $\rho$</td>
<td>Significance level (2-tailed)</td>
</tr>
<tr>
<td>All women</td>
<td>1120</td>
<td>-0.193</td>
</tr>
<tr>
<td>Post-menopausal</td>
<td>961</td>
<td>-0.166</td>
</tr>
<tr>
<td>Pre-menopausal</td>
<td>39</td>
<td>0.180</td>
</tr>
</tbody>
</table>

Table 8.3: Correlation of percentage breast density and glandular volume with age, based on menopause status

It can be seen that the strength of correlation of breast density with age decreases when only the post-menopausal women are considered. It is interesting to note that there are positive correlations between breast density and age in the pre-menopausal group. However, these correlations are not statistically significant and with a sample size of only 39, it is not certain that this relationship would be observed within the screening population as a whole.

The effect of menopause status was further assessed using the null hypothesis that the distribution of breast density (glandular volume and percentage breast density) is the same in pre- and post-menopausal women. Results of the independent samples Mann-Whitney U test are shown below in Table 8.4.

<table>
<thead>
<tr>
<th>Sample size, n</th>
<th>Glandular volume (cm$^3$)</th>
<th>Breast density (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-meno</td>
<td>Post-meno</td>
</tr>
<tr>
<td>Median</td>
<td>69.30</td>
<td>58.46</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>50.07</td>
<td>46.76</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>55.43</td>
<td>42.05</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>112.86</td>
<td>87.55</td>
</tr>
<tr>
<td>Significance</td>
<td>0.013</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Table 8.4: Hypothesis test summary for breast density distribution, based on menopause status

Glandular volume and percentage breast density are significantly different (p < 0.05) in pre- and post-menopausal women and the null hypothesis is rejected. The distribution of glandular volume and percentage breast density within each of the two groups is shown in Figures 8.3 and 8.4 respectively.
To account for potentially confounding factors, independent sample t-tests were carried out to ensure that the distribution of weight and BMI were not significantly different in pre- and post-menopausal women; this was found to be the case. Inevitably, there was a significant difference in age between the two samples.
Multiple linear regression analysis was carried out using age and menopause status as input variables. Age was found to be significantly correlated with both glandular volume ($p < 0.0001$) and percentage breast density ($p = 0.0013$) but menopause status was not.

### 8.2.2 Weight and Body Mass Index (BMI)

The mean weight of the sample was 69.7 kg (s.d. 13.6 kg) and the mean BMI was 26.9 kg m$^{-2}$ (s.d. 5.1 kg m$^{-2}$). Figure 8.5 shows the relationship between glandular volume and weight ($\rho = -0.145$, $p < 0.001$). The correlation with BMI is slightly stronger ($\rho = -0.189$, $p < 0.001$) and is demonstrated graphically in Figure 8.6 using BMI category. The categories are based on the World Health Organisation (WHO) classification [196] shown in Table 8.5 below.

<table>
<thead>
<tr>
<th>BMI</th>
<th>WHO category</th>
<th>Sample size, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 18.5</td>
<td>Underweight</td>
<td>11</td>
</tr>
<tr>
<td>18.5 to 24.99</td>
<td>Normal</td>
<td>445</td>
</tr>
<tr>
<td>25 to 29.99</td>
<td>Overweight</td>
<td>412</td>
</tr>
<tr>
<td>≥ 30</td>
<td>Obese</td>
<td>215</td>
</tr>
</tbody>
</table>

Table 8.5: WHO classification of BMI [196]

![Figure 8.5: Relationship between glandular volume and weight](image)
Figure 8.6: Relationship between glandular volume and BMI category

Figure 8.7 shows the relationship between percentage breast density and weight ($\rho = -0.535$, $p < 0.001$). The correlation with BMI is again stronger ($\rho = -0.594$, $p < 0.001$) and is shown graphically in Figure 8.8 using BMI category.

Figure 8.7: Relationship between percentage breast density and weight
The strength of correlation between percentage breast density with weight and BMI is much greater than that for glandular volume. This is expected, as an increase in weight is typically associated with an increase in body fat. Both weight and BMI are very strongly correlated with total breast volume (weight: $\rho = 0.663$, BMI: $\rho = 0.707$; both significant, $p < 0.001$) and adipose tissue volume (weight: $\rho = 0.670$, BMI: $\rho = 0.718$; both significant, $p < 0.001$).

It is interesting to note that there is a significant negative correlation between weight and BMI with absolute glandular volume, albeit fairly weak. This suggests that the relationship between BMI and percentage breast density cannot be completely explained by an increase in body fat. If the relationship was this straightforward, it is thought that there would be no variation in glandular volume with weight or BMI.

8.2.3 Family history

Family history was defined as having at least one first degree relative (mother or sister) who had breast cancer. 1,121 participants responded to this question and had density data available. Of these, 150 had a family history of breast cancer and 971 did not. The distribution of density data for each group is shown in Figures 8.9 and 8.10 below.
The Mann-Whitney U test was used to evaluate the null hypothesis that the distribution of breast density is the same in women with and without a first degree relative who has had breast cancer. The significance was 0.449 for glandular volume and 0.659 for percentage breast density so the
null hypothesis is retained. Although family history is a risk factor for breast cancer, it has no association with breast density in this cohort.

Because age, weight and BMI were shown to exhibit a significant correlation with breast density and could be confounding factors, it was important that the distribution of these parameters was the same in women with and without family history. Independent sample t-tests showed that this was the case.

8.2.4 Ages at menarche and menopause

The correlation between percentage breast density with age at menarche and age at menopause was found to be very weak with low statistical significance. For this reason, these relationships are not examined further. The correlation between glandular volume and age at menarche was also weak and statistically insignificant. The reason may be that the range of age at menarche is too small to have an effect (mean = 12.84; s.d. 1.65; range 7 – 18 years). However, the correlation with age at menopause was significant for glandular volume ($\rho = 0.110, p < 0.001$).

Mann-Whitney U tests were used to examine whether there was any difference in breast density variables between the categorical groups of ‘early menarche’ versus ‘late menarche’ and ‘early menopause’ versus ‘late menopause’. Early menarche is defined as age < 11 years and late menopause is defined as age > 54 years [66]. There was no significant difference between these groups ($p > 0.05$) for either glandular volume or percentage breast density.

In addition to considering ages of menarche and menopause as separate variables, it is appropriate to consider the relationship between breast density and years of menstruation, since breast density is thought to be related to the rate of breast tissue ageing, which is in turn dependent on hormone exposure [72]. Figures 8.11 and 8.12 show the distribution of glandular volume and percentage breast density respectively with years of menstruation. Some women stated that they had undergone a hysterectomy. If age at hysterectomy was included, this was entered as the age of menopause. This explains why there are some very low values for years of menstruation; the youngest age at hysterectomy was only 21.

Glandular volume increases with years of menstruation, which is expected according to the Pike model. The correlation is statistically significant ($\rho = 0.112, p < 0.001$) although the strength of correlation is slightly lower than those of age, weight and BMI. However, the correlation with percentage breast density was very weak and not significant ($\rho = 0.038, p = 0.240$).
Figure 8.11: Relationship between glandular volume and years of menstruation

Figure 8.12: Relationship between percentage breast density and years of menstruation
8.2.5 Parity versus nulli-parity

1,122 participants responded to this question and had density data available. Of these, 1,021 had children and 101 did not. There were a number of blank responses which could have meant that the woman did not have children but this was not assumed. The second questionnaire addressed this ambiguity by additionally including the question “have you had any children?” The Mann-Whitney U test was used to evaluate the null hypothesis that breast density was the same in women who had children and those who had not. The null hypothesis was retained for glandular volume (p = 0.192) but rejected for breast density (p < 0.05). Percentage breast density was higher in women who had no children (median = 9.24; interquartile range 5.80 - 17.43) than in women who had children (median = 8.67; interquartile range 5.13 - 14.29). Independent samples t-tests showed that age, weight and BMI were not confounding factors.

Late first birth is actually considered a greater risk factor than nulli-parity [66, 67, 72], with ages 35 and 40 both being defined as ‘late’ in the literature. The box-plots in Figures 8.13 and 8.14 show some variation in breast density between women who had no children (n = 101), women who had their first child below the age of 40 (n = 1,011) and women who had their first child at or above the age of 40 (n = 10). The Kruskal-Wallis test was used to evaluate the null hypothesis that breast density was the same in these three groups of women. The results showed that the differences were not significant for glandular volume or percentage breast density, (p > 0.05) so the null hypothesis is retained. It is acknowledged that the sample of women having their first child after the age of 40 (n = 10) is probably too small to permit meaningful statistical analysis.

Figure 8.13: Distribution of glandular volume based on parity and age at first birth
Figure 8.14: Distribution of percentage breast density based on parity and age at first birth

Given the discrepancy in the literature, the Kruskal-Wallis test was also used to evaluate the null hypothesis that breast density was the same in women who had their first child below the age of 35 (n = 990), at or above the age of 35 (n = 31), and women who had no children (n = 101). Again, results indicated no significant difference between these groups (p > 0.05) and the null hypothesis is retained. This also agrees with the finding that age at first child showed no significant correlation with breast density.

8.2.6 Ethnicity

1,253 participants responded to the ethnicity question and had density data available. The ethnicity of the sample is shown in Figure 8.15.

Given that 89.1% are White British, 2.9% are White Irish and 2.3% are Other White, the effect of ethnicity on breast density cannot adequately be assessed using this sample.
8.2.7 Hormone Replacement Therapy (HRT)

Questionnaire 1 did not ask specifically for details about HRT as this information was available in the patient’s notes. Data for both HRT use and breast density was available for 1,120 participants. Of these, 405 had taken HRT during their life and 715 had not. The distribution of density data for each group is shown in Figures 8.16 and 8.17 below.
The Mann-Whitney U Test was used to evaluate the null hypothesis that the distribution of breast density is the same in women irrespective of whether or not they have ever taken HRT. It was found that there was no significant difference in glandular volume ($p = 0.721$) or percentage breast density ($p = 0.202$). The null hypothesis is retained and it is concluded that, within this cohort, the use of HRT at some point during a lifetime does not affect breast density.

However, it is perhaps more important to ascertain whether current HRT use has an effect on breast density. Questionnaire 2 requested information on whether the woman had ever taken HRT and if so, for how many years, were they still taking it and when did they stop.

473 women provided information regarding HRT use. Of these, 250 had taken HRT at some time and 223 had not. However, only 43 were still using HRT. Table 8.6 shows descriptive statistics for the number of years of HRT use, and the number of years since stopping.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of HRT use</td>
<td>7.03</td>
<td>6.00</td>
<td>5.41</td>
<td>10 days</td>
<td>30 years</td>
</tr>
<tr>
<td>Years since stopping HRT</td>
<td>8.53</td>
<td>7.00</td>
<td>5.52</td>
<td>2 months</td>
<td>30 years</td>
</tr>
</tbody>
</table>

Table 8.6: Descriptive statistics for years of HRT use and years since stopping
Comparisons of the breast density distribution for women currently on HRT and not on HRT are shown in Figures 8.18 and 8.19.

Figure 8.18: Distribution of glandular volume based on current HRT use

Figure 8.19: Distribution of percentage breast density based on current HRT use
The null hypothesis is that the distributions of glandular volume and percentage breast density are the same in women who are currently on HRT and those who are not. Results of the Mann-Whitney U Test are shown below in Table 8.7. Age and weight were not found to be confounding factors.

<table>
<thead>
<tr>
<th>Glandular volume (cm³)</th>
<th>Breast density (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Currently on HRT</td>
<td>Not currently on HRT</td>
</tr>
<tr>
<td>Sample size, n</td>
<td>42</td>
<td>196</td>
</tr>
<tr>
<td>Median</td>
<td>78.97</td>
<td>56.83</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>65.02</td>
<td>53.53</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>55.23</td>
<td>40.95</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>102.16</td>
<td>85.49</td>
</tr>
<tr>
<td>Significance</td>
<td>0.002</td>
<td>0.023</td>
</tr>
</tbody>
</table>

Table 8.7: Hypothesis test summary for breast density distribution, based on current HRT use

The null hypothesis is therefore rejected and it is concluded that, within this sample, women currently taking HRT have a significantly higher glandular volume (p < 0.01) and percentage breast density (p < 0.05) than those who are not currently using HRT.

Questionnaire 2 additionally requested information on natural HRT remedies. Of the 472 respondents, 62 had used natural HRT during their lifetime and 8 were still using it. Natural HRT (current or previous use) was found to have no effect on breast density. A summary of the most commonly used natural remedies is shown in Table 8.8.

<table>
<thead>
<tr>
<th>Remedy</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Cohosh</td>
<td>36</td>
</tr>
<tr>
<td>Red Clover</td>
<td>10</td>
</tr>
<tr>
<td>Natural Progesterone</td>
<td>3</td>
</tr>
<tr>
<td>Evening Primrose</td>
<td>3</td>
</tr>
<tr>
<td>“Flash Fighters&quot;</td>
<td>2</td>
</tr>
<tr>
<td>Sage Tea</td>
<td>2</td>
</tr>
<tr>
<td>Soya</td>
<td>2</td>
</tr>
<tr>
<td>Others</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>62</td>
</tr>
</tbody>
</table>

Table 8.8: Use of Natural HRT remedies
8.2.8 Breast cancer

The study participants attended for their screening mammogram between July and December 2007. None were diagnosed with breast cancer during this screening round. The screen-detected cancer rate for the Bolton, Bury and Rochdale screening programme is typically about 0.75% (this is based on the 2010 statistic of 164 cancers detected per 22,000 women screened) [197]. In the study cohort of 1,289 women, about 10 cancers would therefore have been expected.

There were 26 incidences of breast cancer within the sample, detected prior to and after the study period. These are summarised in Table 8.9 below. However, these occurred in only 23 women. Two of the women who had previously had cancer in the left breast had an interval cancer detected in their right breast. One of the women who had previously had cancer in the right breast was diagnosed with cancer in the left breast during the 2010 screen.

<table>
<thead>
<tr>
<th></th>
<th>Diagnosed prior to 2007</th>
<th>Diagnosed during 2010 screen</th>
<th>Interval cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left breast</td>
<td>9</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Right breast</td>
<td>6</td>
<td>5</td>
<td>2 (22/12/09 &amp; 8/2/11)</td>
</tr>
</tbody>
</table>

Table 8.9: Incidence of breast cancer within the study population

The difference in breast density between the breast with cancer and the contralateral breast was examined using paired sample t-tests. The MLO view was used rather than the average for each breast, the reason being that there was a large amount of RCC data missing. Results are shown in Table 8.10. It can be seen that there is no significant difference in glandular volume and percentage breast density between the breast with cancer and the contralateral breast.

<table>
<thead>
<tr>
<th></th>
<th>Glandular volume (cm³)</th>
<th>Breast density (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Difference</td>
<td>95% Confidence Interval</td>
</tr>
<tr>
<td>Breast with cancer -</td>
<td>5.30</td>
<td>-11.65 - 22.24</td>
</tr>
<tr>
<td>contralateral breast</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8.10: Variation in breast density between breast with cancer and contralateral breast

A number of limitations with this analysis are acknowledged: the sample size was very small, reduced further by the fact that several women diagnosed with cancer pre-2007 had undergone
mastectomy. Many had also had lumpectomy or wide local excision which may affect their breast volume and consequently, their breast density.

The Mann-Whitney U-test was used to investigate the null hypothesis that breast density in the contralateral breast of women who had been diagnosed with breast cancer in their lifetime (n = 21) is the same as breast density in women who had not (n = 1,159). The distribution of percentage breast density in these groups of women is shown in Figure 8.20. The null hypothesis is retained and it is concluded that there is no significant difference in glandular volume or percentage breast density between the two samples (p > 0.05). Age, weight and BMI were not found to be confounding factors.

![Figure 8.20: Distribution of percentage breast density based on lifetime breast cancer diagnosis](image)

**8.2.9 Smoking**

There were 477 respondents to Questionnaire 2 with smoking and breast density data available. Of these, 245 had smoked during their lifetime. The Mann-Whitney U-test showed no significant differences in glandular volume (p = 0.147) or percentage breast density (p = 0.217) between those that had smoked and those who had never smoked.

There were 55 respondents who were still smokers. Again, the Mann-Whitney U-test showed no significant difference in percentage breast density between the two groups (p = 0.501). However, there was a significant difference in glandular volume (p < 0.05) with a median of 47.0cm$^3$ (interquartile range 35.8 – 70.0cm$^3$) in the group who were still smokers compared to 61.2cm$^3$ (interquartile range 41.7 – 91.4cm$^3$) in the group who no longer smoked. Age and weight were not confounding factors.
8.2.10 Deodorant use

Of the 477 respondents to Questionnaire 2 with density data available, 461 used an underarm product, whilst only 16 did not. The Mann-Whitney U-test showed no significant differences (p > 0.05) in breast density variables between these two groups. It is acknowledged that the sample size for non-users is very small.

Women were asked which type of product they used: antiperspirant only, deodorant only or a combined product. Most women indicated that they had used more than one type of product so it was impossible to analyse any variations in breast density between women who used different types of product.

8.2.11 Socioeconomic status

The 2007 Indices of Multiple Deprivation (IMD) split by Lower-layer Super Output Area (LSOA) are used as a surrogate for socioeconomic status. This index has been constructed by the Social Disadvantage Research Centre at the University of Oxford [198, 199]. LSOAs are small geographical areas with a population of 1,000 – 3,000 people (typically 1,500). A score is calculated for each of seven domains (income, employment, health, education, housing, crime and living environment); these are weighted and summed to give the overall score. The scores are ranked from 1 – 32,482 with 1 representing the most deprived LSOA in England and 32,482 being the least deprived (although not necessarily the most affluent). It is important to note that the LSOA rank is a relative measure of deprivation so for example, the LSOA ranked 20 is more deprived than the LSOA ranked 40, but not twice as deprived.

The postcode data from Questionnaire 2 was converted to LSOA IMD score, IMD rank and an urban / rural indicator using specialist software [200] courtesy of Terry Child at the Census Dissemination Unit, Mimas, University of Manchester.

Figure 8.21a shows the distribution of LSOA rank for the Questionnaire 2 respondents with breast density data available (n = 471). This is compared to the distribution of the ranks for all LSOAs located in Bolton and Bury [198], Figure 8.21b.

It is interesting to note that the distribution for Bolton and Bury is slightly skewed to the left, indicating a higher proportion of deprived areas within these regions. The median rank is 10,820, which is lower than the mean of 13,567. However, the distribution for the questionnaire respondents is skewed to the right, with a median rank of 19,480 and a mean of 17,126. This suggests that within the population invited for screening, it is the less-deprived women, with perhaps better education and healthcare, who have taken part in the study.
Figure 8.21: Histograms of LSOA rank for (a) Questionnaire 2 respondents and (b) Bolton and Bury

Figure 8.22 shows the relationship between percentage breast density and LSOA rank. There was no significant correlation between the two variables ($\rho = 0.045, p = 0.353$). Additionally, no significant correlation was found between LSOA rank and glandular volume ($\rho = 0.027, p = 0.586$) or LSOA rank and BMI ($\rho = -0.036, p = 0.440$).

Figure 8.22: Relationship between percentage breast density and LSOA rank
Of these 471 respondents, 455 were from urban areas. The remainder were from a *small town*, a *village* or a *hamlet and isolated dwelling*. These data have been pooled and classed as 'rural'. The distribution is shown in Table 8.11 below.

<table>
<thead>
<tr>
<th>Urban / rural indicator</th>
<th>Description of Census Output Area (COA)</th>
<th>Sample size, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Urban with a population of ( \geq 10k ). Wider surrounding area is less sparsely populated</td>
<td>455</td>
</tr>
<tr>
<td>6</td>
<td>Small Town and Fringe area. Wider surrounding area is less sparsely populated</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Village. Wider surrounding area is less sparsely populated</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>Hamlet and Isolated Dwelling. Wider surrounding area is less sparsely populated</td>
<td>3</td>
</tr>
</tbody>
</table>

*Table 8.11: Definitions of urban / rural indicators and the distribution of the sample within each*

The Mann-Whitney U-test was used to assess the null hypothesis that there is no significant difference in glandular volume, percentage breast density or weight between urban and rural populations. The null hypothesis is retained for all variables (\( p > 0.05 \)). However, it is acknowledged that the sample size for the ‘rural’ population is small.

### 8.3 Correlation of risk factors with area-based density

A sample of 268 women from the study population additionally had their mammographic density assessed by a radiologist [186]. The radiologist used a visual analogue scale (Section 3.1.3) to indicate percentage density by area. The distribution of mammographic density in this sample of women is shown in Figure 8.23.

Table 8.12 shows the correlations between mammographic density (%) and volumetric breast density (%) for the left breast with the continuous variables for this sample of 268 women. The histogram of radiologist-assessed breast density, Figure 8.23, is not normally distributed but is skewed to the left, in a similar fashion to that of volumetric breast density (Figure 7.6). For this reason, the Spearman rank correlation coefficient, \( \rho \), is used.
Table 8.12: Correlation of percentage breast density (left breast) by area and by volume with breast cancer risk factors

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Breast density (%) by area</th>
<th>Breast density (%) by volume</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spearman $\rho$</td>
<td>Significance (2-tailed)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.184</td>
<td>0.002</td>
</tr>
<tr>
<td>Weight</td>
<td>-0.466</td>
<td>0.000</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.520</td>
<td>0.000</td>
</tr>
<tr>
<td>Age at menarche</td>
<td>0.140</td>
<td>0.023</td>
</tr>
<tr>
<td>Age at menopause</td>
<td>0.088</td>
<td>0.178</td>
</tr>
<tr>
<td>Years of menstruation</td>
<td>0.018</td>
<td>0.784</td>
</tr>
<tr>
<td>Age at first child</td>
<td>-0.049</td>
<td>0.450</td>
</tr>
</tbody>
</table>

It can be seen that, with the exception of age and age at menarche, volumetric breast density exhibits a stronger correlation with the risk variables. It is reassuring to note that the strength and significance of the correlations with volumetric breast density for this sample generally reflect the findings for the whole study population (Table 8.1). This includes the observation that there was no significant correlation between volumetric density and years of menstruation or age at first child.

8.4 Validity of self-reported data

Assessing the validity of self-reported data has two components: accuracy and reliability [201]. Accuracy is defined as how close the reported measurement is to the true value and reliability is defined as how close initial estimated values are to subsequent estimated values [202].
8.4.1 Accuracy of data

It was thought that weight was the variable most likely to suffer from inaccuracy. A number of studies in the literature found that although errors in self-reported weight were, on average, small within the adult population, the magnitude of the error was directly related to overweight status. Furthermore, women were consistently found to underestimate their weight [203 – 205]. However, a separate study [206] quoted “remarkable accuracy” in self-reported weight, even among the obese, suggesting that measured weight may not be necessary for epidemiological studies.

It was intended that the accuracy of self-reported weight would be assessed by weighing the women who attended at the Bury site, and the ethics committee had granted consent to do this. Weighing would have been infeasible on a screening mobile but the Bury site was larger, with more staff and longer appointment times. Unfortunately, almost all women refused to be weighed and said that they would be unwilling to participate in the study if this was required. Some were not happy with the insinuation that the self-reported data could not be trusted whilst others felt that this should have been stated in the patient information leaflet.

Access to the medical records of participants was not requested as part of the ethics application and therefore published population data has been used to estimate the accuracy of self-reported weight. The mean weight of the study cohort is 69.7 kg (s.d. 13.6 kg). Using BMI as an indicator (which assumes that self-reported height is also accurate), the proportion of women who were found to be underweight, normal, overweight and obese respectively is shown in Table 8.13 alongside data from the Health Survey for England 2007 [207].

The proportion of women within each category in the study cohort is consistent with population data, particularly that for North West women. Although this cannot be taken as evidence that self-reported data are accurate, it does suggest that the low consent rate is not linked to weight and that there is no bias introduced in this sample as a result.

<table>
<thead>
<tr>
<th>BMI category</th>
<th>All women, 2007 Health Survey</th>
<th>North West women, 2007 Health Survey</th>
<th>Study Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underweight (&lt; 18.5)</td>
<td>2.0</td>
<td>2.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Normal (18.5 – 24.99)</td>
<td>41.6</td>
<td>39.0</td>
<td>39.8</td>
</tr>
<tr>
<td>Overweight (25 – 29.99)</td>
<td>32.0</td>
<td>36.0</td>
<td>38.2</td>
</tr>
<tr>
<td>Obese (≥ 30)</td>
<td>24.4</td>
<td>23.0</td>
<td>20.9</td>
</tr>
<tr>
<td>Mean BMI (kgm$^{-2}$)</td>
<td>26.8</td>
<td>26.8</td>
<td>26.9</td>
</tr>
</tbody>
</table>

Table 8.13: Comparison of the distribution and mean BMI between published population data [207] and the study population data
8.4.2 Reliability of data

A comparison of the responses and associated confidence levels between both questionnaires was carried out by Rita Prajapati as a research project in Year 4 of her medical degree. Results and statistical analyses have been taken directly from her report but the interpretation and discussion is my own.

The responses of women who completed both questionnaires were compared using paired t-tests, as shown in Table 8.14.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of responses</th>
<th>Mean (Q1)</th>
<th>Mean (Q2)</th>
<th>Mean difference (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (cm)</td>
<td>446</td>
<td>161.1</td>
<td>160.9</td>
<td>0.1 (-0.2 – 0.3)</td>
<td>0.087</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>413</td>
<td>68.2</td>
<td>68.1</td>
<td>0.2 (-2.3 – 0.6)</td>
<td>0.399</td>
</tr>
<tr>
<td>Age at menarche</td>
<td>460</td>
<td>12.8</td>
<td>12.8</td>
<td>-0.2 (-0.8 – 0.03)</td>
<td>0.432</td>
</tr>
<tr>
<td>Age at menopause</td>
<td>399</td>
<td>48.6</td>
<td>48.2</td>
<td>0.3 (-0.08 – 0.6)</td>
<td>0.056</td>
</tr>
<tr>
<td>Age at first child</td>
<td>480</td>
<td>24.3</td>
<td>23.9</td>
<td>0.4 (-0.01 – 0.7)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Table 8.14: General demographics and results of paired t-tests for women who completed both questionnaires

With the exception of ‘Age at first child’, it is encouraging to see that there was no significant change (p > 0.05) in the reporting of variables between the two questionnaires. The reason for the difference in ‘Age at first child’ is straightforward. In Questionnaire 1, age at first child was requested whereas in Questionnaire 2, age at first pregnancy was requested. Given that women quote their age to the nearest year, there will be a number of instances where this difference is 1 year. Furthermore, some women may not have carried to full-term, in which case the difference may be greater than this.

The difference in the reported age of natural menopause almost achieved statistical significance (p = 0.056). There are two reasons for this: firstly, the question was phrased quite differently in Questionnaire 2 and secondly, the menopause is typically considered as a transitional period of change (perimenopause), making it difficult to quote an exact age. The medical definition of natural menopause states that the ovaries naturally decrease their production of oestrogen and progesterone and there are no menstrual periods for 12 consecutive months [74]. Questionnaire 1 simply asked for ‘age at menopause’, which could include surgical menopause, and many women declared that they had undergone a hysterectomy. Questionnaire 2 was misleading in that
it asked ‘at what age did you start going through the menopause?’ Despite aiming to remove ambiguity in the phrasing of questions, this may have had the opposite effect as it is confusing the menopause with the perimenopause.

8.4.3 Confidence in responses

As discussed in Section 8.1, a Likert scale [188] of confidence was included alongside each question. Results are presented below for Questionnaire 1, taken from the database of the 1,289 women who completed the questionnaire and had breast density data available.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Number of completed entries</th>
<th>% of entries for each level of certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Certain</td>
<td>Quite Sure</td>
</tr>
<tr>
<td>Height</td>
<td>1251</td>
<td>46.8</td>
</tr>
<tr>
<td>Weight</td>
<td>1222</td>
<td>40.1</td>
</tr>
<tr>
<td>Age at first period</td>
<td>1233</td>
<td>42.1</td>
</tr>
<tr>
<td>Age at menopause</td>
<td>1089</td>
<td>30.6</td>
</tr>
<tr>
<td>Age at first child</td>
<td>1157</td>
<td>81.1</td>
</tr>
<tr>
<td>Family history</td>
<td>1262</td>
<td>83.6</td>
</tr>
</tbody>
</table>

Table 8.15: Summary of responses from Questionnaire 1 and associated level of confidence

The strength of correlation between percentage breast density and weight was examined for each level of certainty. Data for ‘not sure’ (n = 5) and ‘don’t know’ (n = 123) were pooled. Responses that had been left blank were excluded from the analysis. The Spearman correlation coefficients were -0.523, -0.529 and -0.441 (all p < 0.01) for the groups ‘certain’, ‘quite sure’ and ‘not sure / don’t know’ respectively.

Tests were also conducted to evaluate the null hypothesis that the mean weight within each level of certainty was the same. Weight was normally distributed and the Levene statistic of 0.783 (significance level = 0.457) confirmed equality of variances between the groups so the one-way ANOVA test was used. The F-statistic was 3.142 (significance level 0.044). The null hypothesis is therefore rejected and it is concluded that there is a difference in mean weight between each level of certainty. This can be seen quite clearly on the means plot in Figure 8.24 below.

It is interesting to note that the confidence in self-reported weight decreases with weight and also that the correlation between weight and breast density is lower in women who are unsure of their weight.
Figure 8.24: Relationship between mean weight and level of certainty in response

Women were more confident in their responses to Questionnaire 2. One possible reason is the fact that they had been asked the question before. Alternatively, this may have been due to the different confidence scale. Questionnaire 2 used ‘certain’, ‘fairly sure’ and ‘don’t know’ whereas Questionnaire 1 used ‘certain’, ‘quite sure’, ‘not sure’ and ‘don’t know’. ‘Quite sure’ is thought to be more ambiguous than ‘fairly sure’ as some may interpret it to indicate a definite response.

8.5 Discussion

Volumetric breast density was found to be associated with a number of risk factors measured on a continuous scale. Of these, glandular volume was found to be significantly correlated ($p < 0.001$) with age, weight, BMI, age at menopause and years of menstruation. However, multiple linear regression analysis revealed that only age ($p < 0.0001$), weight ($p < 0.0005$) and BMI ($p < 0.0001$) retained significance.

Percentage breast density was found to be significantly correlated ($p < 0.001$) with age, weight, and BMI. Weaker and less significant correlations ($p < 0.05$) were observed for age at menarche and age at menopause. However, only age, weight and BMI (all $p < 0.0001$) retained significance in the multiple linear regression model.

These results are consistent with those for mammographic density, which is known to decrease with age [62, 64, 65]. The relationship between mammographic density and age at menarche is uncertain [62]. This also appears to be the case for age at menopause, although when treated as a categorical variable (pre- versus post-menopausal), post-menopausal women exhibit lower
dense area and lower percentage mammographic density (PMD). However, age has been shown to be a confounding factor in such studies [64]. Similar results were observed in this study (Section 8.2.1). Both glandular volume and percentage breast density were found to be significantly lower (p < 0.05) in post-menopausal women but when age and menopause status were used as inputs in multiple linear regression analysis, only age remained significant.

The strength of correlation with weight and BMI is presented for volumetric and mammographic density in Table 8.16. Mammographic density results are taken from a case-control study by Boyd et al [81] where interactive thresholding [60] was used to measure dense area and PMD on the CC view. Only the data from the controls is included. The volumetric breast density data presented below is for the CC view only.

<table>
<thead>
<tr>
<th></th>
<th>PMD (%)</th>
<th>Vol D (%)</th>
<th>Dense Area (cm²)</th>
<th>Gland Vol (cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson r</td>
<td>Spearman ρ</td>
<td>Pearson r</td>
<td>Spearman ρ</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.41</td>
<td>-0.60</td>
<td>-0.10</td>
<td>-0.17</td>
</tr>
<tr>
<td>Significance</td>
<td>p &lt; 0.0001</td>
<td>p &lt; 0.001</td>
<td>p &lt; 0.01</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Weight</td>
<td>-0.38</td>
<td>-0.55</td>
<td>-0.06</td>
<td>-0.14</td>
</tr>
<tr>
<td>Significance</td>
<td>p &lt; 0.0001</td>
<td>p &lt; 0.001</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.001</td>
</tr>
</tbody>
</table>

Table 8.16: Correlation of breast density in the CC view with BMI and weight

For absolute and percentage density, the volumetric parameters demonstrate a stronger correlation with BMI and weight than the area-based parameters. The negative correlation between glandular volume with weight and BMI (discussed in Section 8.2.2) was considered unusual, but this relationship was also observed by Boyd et al for dense area [81]. Interestingly, although the SMF [141] and Volpara™ [208] measures of percentage volumetric density were found to be negatively correlated with BMI, absolute glandular volume exhibited a positive association [141, 208].

No significant correlation was observed between volumetric breast density and age at first birth. Percentage breast density was found to be significantly higher (p < 0.05) in women who were nulli-parous compared to women who had children. Glandular volume was not significantly different between these groups. Woolcott et al [80] obtained similar results, although both dense area and PMD were associated with parity. However, no significant correlation with age at first birth was observed for either variable. Measurement of mammographic density was carried out using interactive thresholding [60] on the CC view. The correlations between parity with mammographic and volumetric density for the CC view are shown in Table 8.17.
<table>
<thead>
<tr>
<th>Parous (≥ 1 child)</th>
<th>Nulli-parous</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean PMD (%)</td>
<td>29.9</td>
<td>34.7</td>
</tr>
<tr>
<td>Median volumetric density (%)</td>
<td>9.6</td>
<td>10.6</td>
</tr>
<tr>
<td>Mean dense area (cm²)</td>
<td>30.0</td>
<td>33.0</td>
</tr>
<tr>
<td>Median glandular volume (cm³)</td>
<td>63.0</td>
<td>61.8</td>
</tr>
</tbody>
</table>

Table 8.17: Correlation of breast density in the CC view with parity

For the other categorical variables examined, only current HRT use and current smoking were found to be associated with volumetric breast density. Using the smaller data sample from Questionnaire 2, it was found that both glandular volume and percentage breast density were significantly higher (p < 0.01, p < 0.05 respectively) in current HRT users compared to those who had discontinued use, or had never used HRT. Use of HRT at some point during a lifetime was not found to affect volumetric breast density. Once again, these results agree very well with those for percentage mammographic density, which has been shown to be higher in current users of HRT [86, 87]. Temporal studies have shown that breast density rapidly decreases following cessation of HRT [65, 86], with such results being observable within just 2 weeks.

Glandular volume was found to be significantly lower (p < 0.05) in those who were current smokers compared to those who had never smoked or who had given up, but there was no significant difference in percentage density between current smokers and non-smokers. Smoking has demonstrated an anti-oestrogenic effect, which is supported by the observation that PMD [106 – 108] and dense area [107] are lower in current post-menopausal smokers. However, smoking is unconfirmed as a risk factor for breast cancer.

Family history was not found to affect volumetric breast density and there remains a lack of evidence as to whether it affects mammographic density [62].

For the sample of women who completed Questionnaire 2, no significant correlation was found between socioeconomic status and glandular volume or percentage breast density. Furthermore, there were no significant differences in volumetric breast density between women living in urban or rural areas, although there were only 16 women in the ‘rural’ dataset. Aitken et al [97] examined the association of mammographic density with deprivation but used a much coarser grading for deprivation score of 1 to 5, as opposed to the LSOA rank [198, 199] of 1 to 32,482. They found that PMD was 6.6% higher in women from the most affluent, compared to most deprived areas but after adjusting for BMI, this dropped to -0.6%. No significant correlation was observed between LSOA rank and BMI for the Questionnaire 2 study cohort. However, Aitken et al [97] found no correlation between deprivation and dense area and no difference in either density variable with urban/rural residence. An observation common to both studies was that the study
population was skewed towards less deprived areas. This is likely to introduce bias into studies evaluating the effect of socioeconomic status on risk.

The effects of ethnicity and deodorant use could not be assessed. 94.3% of the study population were ‘White’ (British / Irish / Other) and 96.6% used underarm products, making meaningful statistical analysis impossible.

The same comment can be made for the attempt made to examine the relationship between volumetric breast density and breast cancer risk. There were only a small number of cancers in the study sample (n = 26) and none were diagnosed during the screening round over which the study took place. Analysis was additionally limited by the fact that these occurred in 23 women, of whom 2 had a mastectomy and a further 7 had surgery which would have affected their breast volume and density. Consequently, no significant differences in volumetric breast density were observed between women who had been diagnosed with breast cancer in their lifetime and those who had not.

This lack of association is disappointing but somewhat expected. With an NHSBSP cancer detection rate of only 7.8 per 1,000 women screened [40], prospective studies investigating the relationship between breast density and breast cancer risk would need to recruit extremely large numbers of women to give statistical confidence in the results. Alternatively, a retrospective case-control study could be carried out. The current Manchester Method is unsuitable for such studies because of the requirement for the stepwedge and magnification markers to be imaged alongside the breast. However, it is anticipated that when the method is adapted for digital mammography, retrospective analysis should be possible. This is discussed further in Chapter 9.

Although volumetric breast density using the Manchester Method was not found to be associated with breast cancer risk in this particular study, it did exhibit strong and significant correlations with a number of other risk factors. The results of this study agree well with those in the literature examining the relationship between mammographic density and age [62, 64, 65], menopause status [64], weight and BMI [81], parity [80], HRT use [86, 87], smoking [107] and socioeconomic status [97]. The correlation between breast density with weight and BMI was stronger for volumetric density than mammographic density, based on radiologist assessment (Section 8.3) and interactive thresholding [81]. These are certainly encouraging findings which are suggestive of an association between volumetric breast density and breast cancer risk, especially when taken in combination with the strong relationship between volumetric and area-based density measurements demonstrated in Chapter 6.

It is believed that this is the largest study to date (n = 1,200) evaluating the correlations between volumetric breast density and a comprehensive set of breast cancer risk factors. Other studies using SMF [141] (n = 590) and Volpara™ [208] (n = 370) found that percentage volumetric density
was strongly related to age, BMI, nulliparity, age at first birth and menopausal status. However, the strength of correlation was the same, or weaker, than with mammographic density measured by interactive thresholding. Quantra™ has been found to be significantly correlated with age, menopausal status, family history and HRT use [209], but the population size was again smaller (n = 320) than in this study.

This feasibility study has acted as the pilot for a much larger study known as PROCAS [112, 113] (Predicting Risk Of Cancer At Screening) which is currently underway in Manchester. One of the aims of this study is to determine which method of breast density measurement demonstrates the strongest association with breast cancer risk. The five methods being used are the visual analogue scale, Cumulus [60], Quantra™ [153], Volpara™ [154] and the Manchester Method. Further details on the PROCAS study are presented in the final chapter.
9. Adaptation of the Manchester Method to Full Field Digital Mammography

In December 2007, the Department of Health (DH) published the *NHS Cancer Reform Strategy* [210]. In order to diagnose breast cancer earlier, it was proposed that the invited age range for screening be gradually extended to 47 – 73 years, with programme expansion completed by 2012. Direct digital mammography would also be introduced over this period, as this had been shown to offer improved sensitivity in women under 50 compared to film-screen mammography [160]. Breast screening centres were required to have at least one digital mammography unit by 2010.

Unfortunately, there has been a delay in meeting these requirements. The age expansion will not be completed until 2016. At July 2011, only 85% of UK breast screening centres had at least one digital mammography unit [10]. Nevertheless, digital mammography is rapidly replacing film-screen mammography.

The challenges of digital mammography were discussed in Section 3.3. Although area-based measures of mammographic density have shown a strong association with breast cancer risk when applied to film-screen mammograms [18, 21, 116], this relationship has not yet been established for digital mammography images. Although one study described the use of interactive thresholding on raw images [119], it is expected that the processed images would more commonly be used for analysis, as this is the image format that the radiologist would use for diagnosis. Mammographic density has been found to be lower on these images than film-screen images when assessed visually [115] and using Cumulus [161]. This raises the concern that if breast density is underestimated then breast cancer risk may also be underestimated.

Techniques for measuring volumetric breast density on raw digital mammography images include Quantra™ [153] and Volpara™ [154], which are extensions of the Standard Mammographic Form (SMF) [134]. Measurements of volumetric density using these techniques have shown strong correlations with Cumulus v4 [118, 119]. Calibration methods also exist, such as Cumulus V [158, 159] and Single X-ray Absorptiometry (SXA) [157]. However, there is no literature to date on the association of volumetric breast density with breast cancer risk for any of these techniques.

Although this is an active area of research, further investigations are required. This chapter describes a proposal for how the Manchester Method will be adapted for digital mammography, including some preliminary calibration data and results on detector stability.
9.1 Digital detector technology

Full Field Digital Mammography (FFDM) encompasses a number of detector technologies, which are briefly described below.

9.1.1 Computed Radiography (CR)

Computed Radiography (CR) utilises the physical principle of photostimulable phosphor luminescence. The main advantage of this technology is that it can be used with existing analogue mammography equipment by replacing film-screen cassettes and processors with phosphor plate cassettes and CR readers.

However, a major limitation is that for a given level of image quality, CR requires higher doses than film-screen and substantially higher doses than integrated digital solutions [211, 212]. On 26 May 2010, the DH Advisory Committee on Breast Cancer Screening decided that direct digital technology was the preferred option for the NHSBSP [10]; the same recommendation was made for screening mammography in Australia in July 2012 [211]. CR systems will therefore represent a very small proportion of all FFDM units.

CR exhibits non-linear relationships between dose to the image plate, detector dose indicator (also referred to as Exposure Index) and grey level value. This is complicated further by the fact that these relationships vary between vendors and also within vendors, depending on the image processing algorithm selected. It is not simply a case of choosing a raw or processed image. An example is given in Figure 9.1 for the Fuji Profect CR system, which shows some of the image processing algorithms that can be selected for Quality Control (QC) testing. There are in fact four pages of algorithms for QC, with additional algorithms for clinical images.

![Figure 9.1: Screen-shot of the Fuji Profect CR system, illustrating the extensive range of image processing algorithms available for QC. There are additional options for clinical images](image-url)
Modifying the Manchester Method to account for these complexities would involve a significant amount of work. Given that CR will not be widely adopted, it is difficult to justify such an undertaking.

9.1.2 Integrated Digital Detectors (DR)

Integrated digital technology includes flat panel detectors and scanning photon-counting systems. The latter is known as the MicroDose system and was designed by Sectra Medical Systems. It employs multiple scanning slit technology with 28 fan beams, where pre- and post-breast collimation is used to reduce scatter to the extent that a grid is not required. This is used in combination with the novel technique of single photon-counting with energy discrimination, which results in the rejection of scattered photons and electronic noise. Consequently, doses are substantially lower than for film-screen mammography [211]. Sectra has recently developed spectral mammography which has applications in breast composition analysis [213]. The Manchester Method will not be adapted for photon-counting technology.

Flat panel detectors may be indirect or direct. Indirect solutions are based on amorphous silicon (a-Si) flat panel detectors coupled to a CsI scintillator. There are therefore two energy conversions: x-ray photons to light in the scintillator and light to electronic signal in the a-Si photodiode array. Direct flat panel detectors utilise amorphous selenium (a-Se) technology which converts x-rays directly to electronic signals. The single energy conversion means that these detectors should exhibit superior detective quantum efficiency (DQE). A desirable property of flat panel detectors is that there is a linear relationship between grey level value (mean pixel value, MPV) measured on the raw (“for processing”) images and dose to the detector. This is shown in Figure 9.2 for the GE Senographe Essential, which uses an a-Si detector and the Hologic Selenia Dimensions, which uses an a-Se detector. The data were collected during medical physics QC testing.

This relationship is certainly advantageous for calibration techniques because it has been shown that, in principle, the detector itself can act as the calibration device [151]. Thus, the stepwedge used in the current Manchester Method should no longer be required. This approach is investigated in the next section.
9.2 The Manchester Method for Full Field Digital Mammography

Preliminary calibration data have been collected on the GE Essential and Hologic Lorad Selenia mammography units, using the raw images. As described in Section 4.4, this involves making multiple x-ray exposures of breast tissue equivalent phantoms encompassing a range of thicknesses and compositions. The materials used are epoxy resin based tissue substitutes AP6 and WT1 which simulate fatty and glandular tissue respectively [142].

Data were collected using manual factors which covered a range of kV, target / filter (T/F) combination and mAs. When using film, which is expensive, calibration data were collected under automatic exposure control in order to best represent the exposure factors that would be
selected clinically. However, when using digital mammography, there are no cost implications with the number of x-ray images that can be acquired, so the effect of kV and T/F can be investigated. The GE unit uses molybdenum (Mo) and rhodium (Rh) as the target and filter materials. The possible T/F combinations are Mo/Mo, Mo/Rh and Rh/Rh. The Hologic unit uses a tungsten (W) filter with rhodium (Rh) and silver (Ag) filters. The possible T/F combinations are W/Rh and W/Ag but W/Ag has a hard beam quality and is rarely selected clinically.

Calibration data for the GE Essential were collected using the same exposure factors as Kaufhold et al, who used a GE Senographe 2000D in their study [151]. On the Hologic Lorad Selenia, an initial exposure was made under automatic exposure control for every combination of phantom thickness and composition. This was used to inform the range of manual factors that were set.

Figures 9.3 and 9.4 show data collected on the Hologic Lorad Selenia and GE Essential units respectively. Mean pixel value (MPV) is defined as the average grey level measured in a Region of Interest (ROI), which was selected to have a size of 10x10mm and was positioned in the centre of the phantom. Although consistency was achieved, the ROI position is not considered too critical because even raw digital images have had a “flat-field” correction applied which removes variations in signal intensity caused by the anode-heel and inverse square law effects, thereby ensuring a constant grey level across the image. Rather than normalising to a standard mAs (product of tube current and time), which was the approach taken by Kaufhold et al (110mAs) [151] and Malkov et al (100mAs) [157], the variable (MPV/mAs) was plotted against glandular tissue thickness. This resulted in an exponential relationship; hence the natural log, ln (MPV/mAs), is shown on the graphs below.

![Graph](image)

Figure 9.3: Relationship between glandular thickness and ln(MPV/mAs) on the Hologic Selenia; the legend shows total breast thickness and kV/T/F and the trendline shows the relationship between these variables and the strength of fit.
Figure 9.4: Relationship between glandular thickness and ln(MPV/mAs) on the GE Essential for a range of kV and T/F combinations; the legend shows total breast thickness and the trendline shows the relationship between these variables and the strength of fit.
On each of the figures above, a linear trend line has been fitted to the data for a total phantom thickness of 40mm. The $R^2$ value exceeds 0.99 in all cases, demonstrating excellent strength of fit. However, the different gradient (m) and intercept (c) values, summarised in Table 9.1, indicate that the relationship between glandular tissue thickness and ln (MPV/mAs) is dependent upon kV and T/F combination.

<table>
<thead>
<tr>
<th>GE Essential, Mo/Mo</th>
<th>Hologic Selenia, W/Rh</th>
</tr>
</thead>
<tbody>
<tr>
<td>kV</td>
<td>m</td>
</tr>
<tr>
<td>25</td>
<td>-25.79</td>
</tr>
<tr>
<td>26</td>
<td>-26.55</td>
</tr>
<tr>
<td>27</td>
<td>-27.88</td>
</tr>
<tr>
<td>28</td>
<td>-28.98</td>
</tr>
</tbody>
</table>

Table 9.1: Gradient and intercept of the linear fits to the data for 40mm phantom thickness, demonstrating the effect of kV and T/F

This implies that it is necessary to collect calibration data for a range of kV and target / filter combinations. To prevent the volume of calibration data from becoming unmanageable, exposures made under automatic exposure control (AEC) could be used to inform the combinations required for each breast thickness, as shown in Table 9.2. The AEC modes are known as *Std Auto* for GE units and *Auto Filter* for Hologic units. For both modes, kV and T/F are selected automatically and the exposure terminates at an mAs corresponding to a pre-determined level of image signal to noise ratio (SNR).

<table>
<thead>
<tr>
<th>Total thickness (mm)</th>
<th>Gland thickness (mm)</th>
<th>GE Essential, Std Auto</th>
<th>Hologic Selenia, Auto Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>kV</td>
<td>T/F</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>25</td>
<td>Mo/Mo</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>26</td>
<td>Mo/Mo</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>26</td>
<td>Mo/Rh</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
<td>29</td>
<td>Rh/Rh</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>29</td>
<td>Rh/Rh</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>29</td>
<td>Rh/Rh</td>
</tr>
<tr>
<td>60</td>
<td>30</td>
<td>29</td>
<td>Rh/Rh</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>30</td>
<td>Rh/Rh</td>
</tr>
</tbody>
</table>

Table 9.2: Factors selected under automatic exposure control which could be used to inform the range of kV and T/F permutations required for calibration data
Digital mammography units use in-built look-up tables to determine the factors selected under AEC. The Hologic unit selects kV based solely on the system-indicated breast thickness. A high kV (e.g. 30kV) would never be selected for a 2cm compressed breast, regardless of its density, so calibration data would only need to be generated over a small range of kVs for each breast thickness. The GE unit selects exposure factors based on the system-indicated breast thickness and a short pre-exposure which gives an indication of composition. For a given breast thickness, both the kV and T/F combination can vary depending on the breast density. However, the Rh/Rh combination would never be selected clinically for a 2cm breast and Mo/Mo would never be selected for a 6cm breast so there would be no need to collect calibration data for certain thickness and T/F permutations.

Rather than generating a single calibration curve relating grey level value, glandular tissue thickness and total breast thickness (Figure 4.20), it is proposed that numerous calibration curves are created in order to account for the effects of kV and T/F. Additionally, ln (MPV/mAs) should be used instead of grey level. The relationship between glandular tissue thickness ($x_g$) and ln (MPV/mAs) is given by Equation 9.1:

$$x_g = m \left( x_b, kV, T / F \right) \ln \left( \frac{MPV}{mAs} \right) + c \left( x_b, kV, T / F \right) \quad (9.1)$$

Both the gradient (m) and intercept (c) are functions of total breast thickness ($x_b$), kV and target/filter combination.

Fortunately, all digital images have an associated DICOM (Digital Imaging Communications in Medicine) header [152] which contains the exposure factors under which the image was acquired. Figure 9.5 shows an example for the GE Essential. This could be utilised to reference the appropriate set of calibration data for each image.

![Figure 9.5: DICOM header for GE Essential. Relevant DICOM tags are highlighted](image)
9.3 Detector stability

Before applying such an approach in a clinical study, it is necessary to confirm the suitability of the detector as the calibration device. Kerrison et al [156] investigated whether the initial calibration of GE Essential units could be used during routine mammography for extended periods without the need to recalibrate. There were two components to their study. The first was to compare the calibrations on two units over a period of time to evaluate whether the calibration data sets differed in absolute or relative terms. The second was to monitor the results of the daily quality control test (expressed as MPV/mAs) for five units, to quantify the performance stability and determine whether such measurements could be used to inform the need for recalibration.

Three sets of calibration data were compared from two GE Essential units; one was the unit discussed in Section 9.2, for which calibration data were collected in November 2009. The other unit was located in a different room at the same breast screening centre and calibration was carried out in January 2012 and repeated in March 2012. Very little difference was observed between the lines of fit to the calibration data, with average differences of <1% in gradient and <2% in the intercept values. However, the absolute values of MPV/mAs for a given breast thickness, composition, kV and T/F were different between the units and over time. This is somewhat expected, based on the results of QC tests. Although system performance should be similar for a given manufacturer and model of mammography unit, slight variations in detector response between equipment are normal, as illustrated in Figure 9.6 which shows a comparison of six GE Essential units tested in 2011 and six Hologic Selenia Dimensions units tested in 2012. The Hologic units were new at this time, which may explain the lower variation.

![Figure 9.6: Comparison of the detector response for 6 GE Essential units (left) and 6 Hologic Selenia Dimensions units (right)](image)
It is therefore necessary to collect an initial calibration data set for every mammography unit. However, Kerrison et al [156] proposed that it should be possible to account for temporal changes in system performance by scaling the original calibration data by the ratio of \((\text{MPV/mAs})_{\text{current}}\) to \((\text{MPV/mAs})_{\text{initial}}\), when MPV/mAs is obtained under standardised conditions. This was found to predict glandular tissue thickness to within 1mm of the value found using the original calibration data.

The analysis of daily quality control results revealed that detector performance was extremely stable over long periods of time for static units. For example, the MPV/mAs for one unit remained within ± 5% of the baseline value for a period of 244 days (the total duration of data available). Greater variation was observed for mobile units, which is thought to be related to changes in temperature and humidity, as these are more difficult to control in a mobile environment. Systematic deviations in MPV/mAs could generally be attributed to a GE service visit, with significant changes being the result of equipment changes, such as recalibrating or replacing the detector [156].

**9.4 Discussion**

Preliminary work has been carried out towards adapting the Manchester Method to make it suitable for digital mammography units with integrated flat panel detectors. The advantages of digital technology include the linear detector response, the “flat-field” correction of images and the DICOM header which contains pertinent information associated with each image. The main disadvantage is that the method must be applied to the raw images, which are not routinely sent to the Picture Archiving and Communications System (PACS) due to cost and storage limitations.

Section 9.2 shows that, in theory, the detector can act as the calibration device and the stepwedge should no longer be required. Glandular tissue thickness is linearly related to \(\ln \left(\text{MPV/mAs}\right)\) but the gradient and intercept of the lines of fit are dependent upon total breast thickness, kV and target / filter combination. It is therefore necessary to generate a number of calibration curves and use the exposure factors contained in the DICOM header to reference the appropriate curve for each image.

A potential concern associated with using the detector as the calibration device is the temporal stability in performance. Kerrison et al [156] have shown that the MPV/mAs measured during daily QC tests remains sufficiently constant over long periods of time (exceeding 244 days) and that significant deviations from the baseline results are typically explained by a service visit. Monitoring of routine QC results could therefore be used to determine when recalibration is required. Interestingly, Kerrison et al [156] found that it may not be necessary to generate full calibration data sets on such occasions, but that the initial calibration data could be scaled based on differences in MPV/mAs obtained under standardised conditions. Their investigation
showed that this would result in errors in glandular tissue thickness measurement of up to 1mm, compared to using the original data. Stability in the daily MPV/mAs values was defined as a variation of less than ± 5% from baseline, which was estimated to equate to errors in glandular tissue thickness of less than 2mm [156]. Further work is required to determine how this would affect glandular volume and percentage breast density. Additionally, their study included only GE modalities, which use indirect flat panel detectors (a-Si). An investigation of the stability characteristics of direct flat panel detectors (a-Se) is therefore required.

The Manchester Method for digital mammography has not yet been applied clinically. The use of the detector as the calibration device means that there is potential to apply the technique retrospectively, but only if an alternative method for measuring compressed breast thickness is developed. This must account for paddle tilt and must not compromise the level of accuracy achieved by the magnification markers, as this is a major advantage of the current method. The system-indicated value does not meet these criteria. However, Figure 4.7 shows that there are well defined relationships between measured breast thickness, system-indicated breast thickness and applied compression force, which could form the basis of a breast thickness calibration data set. These values would be available in the DICOM header. A similar approach [149] is employed in Cumulus V [158, 159].

This represents an area of future work since techniques that can be used retrospectively can be applied in case-control studies, which represent the best method for determining the strength of association between breast density and breast cancer risk. Quantra™ [153] and Volpara™ [154] are superior to the current Manchester Method in this respect. However, they use the system-indicated breast thickness (with a fixed paddle tilt assumed by Volpara™) which will lead to uncertainties in the accuracy of their volumetric breast density measurements; this was found to be the case when Volpara™ was used on a mammography unit (Hologic Selenia) where the compression paddle was specially designed to tilt [208].
10. Conclusions and Future Work

This chapter re-visits the original aims of this PhD and describes how each has been met. This is followed by a summary of the contributions to knowledge and proposals for future work.

10.1 Meeting the aims and objectives

The aims and objectives were originally defined in Chapter 1; the same headings are used below.

10.1.1 Extend the Manchester Method for volumetric breast density measurement

The principal achievement is the design of a new stepwedge, which has enabled the Manchester Method to be used in the screening programme for the first time. The new aluminium stepwedge is more compact and lighter than the original PTFE stepwedge, as lead lining is not required. Additionally, the improved stepwedge encompasses the range of optical densities encountered clinically for all breast thicknesses and densities. The mechanism for attaching the stepwedge to the breast support platform represents a further design improvement; it remains securely in place during MLO views and can easily be flipped out of the field of view on the rare occasions that it will not fit alongside the breast.

The new marker sheets are more robust and can be constructed with a precision that is highly repeatable. The method of breast thickness measurement has been shown to be extremely accurate and accounts for paddle tilt.

A realistic model of thickness in the breast margin has been incorporated for the first time, resulting in improved accuracy of breast volume, glandular volume and percentage breast density measurements.

10.1.2 Validate the new method

The Manchester Method is reliable, having demonstrated excellent intra- and inter-observer agreement, far superior to the agreement between two radiologists performing visual assessment of mammographic density. Although the technique is semi-automated, each of the processes requiring operator input generates reproducible results. The greatest operator dependence arises from marking up the stepwedge but the average absolute difference in percentage breast density between observers was only 0.26% (CC view) and 0.38% (MLO view).

Glandular volume and percentage breast density have been shown to be highly correlated with visually assessed mammographic density. Mammographic density measured in this way is
strongly associated with breast cancer risk, suggesting that volumetric breast density measured using the Manchester Method should also be strongly related to risk. Four categories of risk were proposed for percentage breast density and the distribution of data within each category was found to be very similar to that for mammographic density categorised using the BI-RADS classification.

10.1.3 Evaluate the feasibility of using the improved method in the breast screening programme

Use of the Manchester Method in the screening programme was deemed to be viable. Feedback from the radiographers and radiologists involved in the study indicated that the method did not impede the screening examination and did not interfere with the reading of mammograms. However, there were a number of method failures attributed to radiographer error, for example, placing an address label over the stepwedge or not including the marker sheet on the compression paddle. It is thought that a large proportion of these could be eliminated in future studies with better staff education and supervision in the early stages.

The Manchester Method generated volumetric breast density measurements for all women taking part in the study (excluding those where the stepwedge and / or markers were missing), indicating that it can cope with the range of breast thicknesses and compositions encountered in the screening population. However, confidence in the measurements of breast density was limited in approximately 1% of cases, as the grey levels within the stepwedge did not cover the full range of grey levels expected within correctly exposed breast tissue. Unfortunately, these were the women with the thickest, densest breasts.

10.1.4 Obtain information on the distribution of volumetric breast density in the screening population

This was the first time the Manchester Method had been applied to the screening population. Breast thickness was found to be normally distributed within the study cohort but breast area, volume, glandular volume and percentage breast density were positively skewed. Percentage breast density exhibited a negative relationship with breast thickness and breast volume, but interestingly, glandular volume remained independent, although the variance was large.

The median percentage breast density was 8.4% (interquartile range 4.9 – 14.2%). There was no significant difference between the left and right breasts and the correlation between them was stronger than reported for other volumetric techniques. The difference between the CC and MLO view was significant, but closely correlated. On average, the CC view was found to be denser than the MLO view, an observation also made using radiologist visual-assessment [116].
reason, correlations between volumetric breast density and risk factors were examined using LCC and LMLO data.

The median glandular volume was 60.1cm$^3$ (interquartile range 42.2 – 86.3cm$^3$) and was not significantly different between left / right breasts and CC / MLO views.

10.1.5 Correlate absolute glandular volume and percentage breast density with established breast cancer risk factors and risk

For risk factors measured on a continuous scale, glandular volume was found to be significantly correlated with age, weight, BMI, age at menopause and years of menstruation. Percentage breast density was found to be significantly correlated with age, weight, BMI, age at menarche and age at menopause. However, for both variables, multiple linear regression analysis revealed that only age, weight and BMI retained significance. For risk factors measured on a categorical scale, both volumetric density measures were higher in women who were currently on HRT, but previous use of HRT had no effect. Percentage breast density was found to be higher in women who were parous and glandular volume was higher in women who were current smokers.

These results are consistent with those for mammographic density reported in the literature. The correlation between breast density with weight and BMI was stronger for volumetric density than mammographic density, measured by visual assessment and interactive thresholding [81]. These are certainly encouraging findings which are suggestive of an association between volumetric breast density and breast cancer risk. Unfortunately, no correlation with breast cancer risk was observed in the feasibility study as there were only a small number of cancers in the cohort (n = 26). In order to establish such a relationship, it is necessary to apply the Manchester Method to a case-control study or carry out a prospective study with a much larger population, such as the PROCAS study [112, 113] described below.

It is believed that this is the largest study to date (n = 1,200) evaluating the correlations between volumetric breast density and a comprehensive set of breast cancer risk factors. Studies using SMF [141] and Volpara™ [208] had smaller study populations and the correlations between risk factors and volumetric density were no greater than with mammographic density measured by interactive thresholding.

10.1.6 Adapt the method to make it suitable for use in digital mammography

Preliminary work has shown that an integrated digital detector can act as the calibration device for the Manchester Method, so the stepwedge should no longer be required. It is necessary to generate a number of calibration curves and use the exposure factors contained in the DICOM header to reference the appropriate curve for each image. Furthermore, concerns about the
temporal stability of the detector mean that routine quality control results should be monitored to
determine when the initial calibration data should be corrected, or if new calibration data should be
collected.

10.1.7 Adapt the method to measure density by area and examine the association with
visually assessed density

Mammographic density has been criticised because projected dense area is not a true reflection
of the amount of dense tissue within the breast. The Manchester Method was extended to derive
area-based measures of density by defining ‘dense pixels’ based on the percentage of glandular
tissue within the column of breast tissue projected on to that pixel.

A radiologist-defined dense pixel was found to be one in which the percentage of glandular tissue
was at least 10 to 20% of the total thickness of the compressed breast at that point. However, no
single threshold demonstrated perfect agreement with either of the two radiologists in the study so
it is not proposed that the Manchester Method be used in this way to provide semi-automated
measures of percentage dense area, or to use the threshold value to convert percentage dense
area to percentage dense volume.

10.2 Contributions to knowledge

The major contributions to knowledge are summarised as follows:

- A new stepwedge has been designed, which has enabled the Manchester Method to be
  used in the screening programme for the first time.
- A realistic breast edge model has been incorporated into the method, which improves the
  accuracy of volumetric breast density measurements.
- Descriptive statistics about breast thickness, area and volume, as well as glandular tissue
  volume and percentage breast density have been computed for the screening population.
  The relationships within and between these measures have been evaluated.
- Correlations between breast cancer risk factors and volumetric breast density have been
determined and were found to be at least as strong as those using mammographic
density.
- Strong and significant relationships were found to exist between volumetric breast density
  and visually assessed mammographic density, when the latter was treated as a
  continuous and categorical variable. This suggests that volumetric breast density should
  be associated with breast cancer risk.
- Strategies for successful study design and optimisation of participation rate were informed
  by the feasibility study and have been applied to the PROCAS study.
Investigations into adapting the Manchester Method for digital mammography suggest that integrated digital detectors can be used as the calibration device, provided that the temporal stability of the detector is monitored.

The minor contributions include:

- The design of more robust marker sheets.
- The design of an improved mechanism for attaching the stepwedge to the breast support platform.
- An insight into how a radiologist classifies a dense pixel.
- The determination of correlations between volumetric breast density and lesser-known risk factors, such as smoking and socioeconomic status.

10.3 Future work

Proposals for further development and validation of the Manchester Method are presented in the following sections.

10.3.1 Improvements and extensions to the Manchester Method

Although the elliptical model of breast thickness in the breast margin is considered realistic and represents a major improvement over the rectangular model previously applied [37, 163], there were a small number of cases where the glandular thickness map indicated that glandular volume had been overestimated in this region. An example is shown in Figure 10.1. The breast margin appears as a bright band and is separated from the central compressed breast region by a black line, although this is not continuous.

Figure 10.1: Example of overestimating glandular thickness within the breast margin
An overestimate of breast density has been shown to result from an underestimate of tissue thickness. This may be attributed to an inappropriate model of breast thickness within the margin, or incorrect definition of the inner edge of the breast margin, which will affect the horizontal and vertical radii of the ellipse.

The breast margin is a region of low breast thickness and is known to be composed of predominantly adipose tissue. However, it is felt that an attempt should be made to quantify the volume of dense tissue within the margin and therefore further work is required.

Processed digital mammography images may aid with the accurate determination of the inner and outer breast margin. This is demonstrated in the images in Figure 10.2 where the compressed breast region is clearly discernible from the breast margin. However, this appearance is actually considered undesirable by radiologists who prefer to visualise the same level of grey level intensity across the whole breast; tissue thickness equalisation algorithms have therefore been developed by the manufacturers of digital mammography units.

![Figure 10.2: Digital mammography images with poor tissue thickness equalisation; the breast is clearly separated into the compressed breast region and the margin](image)

Tomosynthesis is an advanced application of digital mammography where a number of projection images are required over an angular range and the breast is reconstructed as a series of slices. There may be potential to use tomosynthesis slices to detect the edges of the breast margin and to give an indication of the change in breast thickness within this region.

Further extensions to the Manchester Method include an option for the user to manually segment the breast if automatic segmentation fails. Even though the current implementation of the method has been found to be reliable, a fully automated technique would be desirable and this would require automatic segmentation of the pectoral muscle and automated detection of steps within the stepwedge, although the latter would no longer be required for digital mammography.
10.3.2 Application to Full Field Digital Mammography

Preliminary work has been undertaken and has shown that the integrated detector may act as the calibration device, as long as routine quality control results are monitored to assess detector stability. Calibration data have been collected for a-Si (indirect conversion) and a-Se (direct conversion) detectors but this has not yet been incorporated into the analysis software and therefore the method has not been applied clinically. It is anticipated that the digital Manchester Method will be included in the PROCAS study [112, 113] but further developments are required.

Using the detector as the calibration device means that the Manchester Method has the potential to be applied retrospectively, for example in case-control studies. However, an alternative technique for accurately measuring breast thickness would be required.

10.3.3 Three-dimensional (3D) techniques

Three-dimensional breast imaging techniques include magnetic resonance imaging (MRI), dedicated breast computed tomography (DBCT) and tomosynthesis. All have the advantage of displaying the breast as a series of slices, thereby removing the superimposition of structures associated with conventional mammography where the image is a two-dimensional projection of the breast.

The accuracy of glandular tissue volume and percentage breast density measurements generated by the Manchester Method has not yet been investigated. It is recommended that this is carried out by comparing the results to those from a 3D technique. Both MRI and DBCT have been shown to be suitable for this purpose [153, 154, 158]. However, they are considered to be diagnostic, rather than screening, tools and there is no suggestion to use them for routine breast density assessment as it would be difficult to collect sufficient data to determine the relationship between density and risk.

Tomosynthesis is not currently used for screening mammography but investigations are underway to determine whether it could be employed in this capacity. It would therefore be worthwhile to develop a volumetric breast density method which could be applied to tomosynthesis images.

10.3.3 The PROCAS Study

The aim of the PROCAS study [112, 113] is to determine whether the routine assessment of breast cancer risk is feasible within the NHSBSP and if women would wish to know the outcome of their risk assessment. Individualised risk prediction models have the potential to be used as the basis for tailoring the screening interval and for recommending risk-reducing interventions such as lifestyle changes and the use of oestrogen receptor modulators, such as tamoxifen.
New prediction models will be developed which incorporate additional risk factors such as breast density. Breast density is being measured using five methods: the visual analogue scale, Cumulus [60], Quantra™ [153], Volpara™ [154] and the Manchester Method. The study will therefore enable determination of the relationship of volumetric breast density measured by the Manchester Method and breast cancer risk. It will also indicate which technique demonstrates the strongest association with risk.

Furthermore, measurements made by the Manchester Method can be compared to those from other volumetric techniques on the same sample of mammograms. The proposed study time is 6 years, which incorporates two screening rounds. The temporal consistency of the density measurement techniques can therefore be assessed.

10.4 Conclusion

Breast density is an important independent risk factor for breast cancer and has been shown to improve the accuracy of risk prediction models. It is of particular interest because it can be reduced by interventional methods although it is not yet known if a reduction in density carries a corresponding reduction in risk.

Prior to commencing this PhD, methods for measuring breast density were largely subjective and estimated mammographic density, which represents the projected area of dense tissue on a mammogram. The scientific validity of this approach has been heavily criticised. Objective techniques capable of measuring the volume of dense tissue were in varying stages of development, with SMF [134] being the most established. However, volumetric breast density had not demonstrated an association with breast cancer risk; this remains the case.

This PhD has resulted in the development of a highly reproducible technique which measures glandular tissue volume and percentage breast density. It is feasible to use in the breast screening programme, where measurements of volumetric breast density were found to be related to a number of breast cancer risk factors, with the strength of association exceeding that of visually assessed mammographic density. Furthermore, both glandular volume and percentage breast density were found to be strongly correlated with visually assessed mammographic density. These are promising findings which are suggestive of an association between volumetric breast density and breast cancer risk. However, further work is required to establish a direct relationship with risk.

Research into breast density continues to gather momentum, particularly in light of the independent review into the efficacy of the UK breast screening programme. Although results have not yet been published, much interest has already been generated into tailoring the
screening interval according to individual risk. Breast density will be an integral component in these risk assessments.
References


31 Kopans DB. Basic physics and doubts about the relationship between mammographically determined tissue density and breast cancer risk. Radiology 2008; 246: 348-353
32 Yaffe M, Boyd N. Quantifying breast cancer risk from mammograms. CARS 2002; 5: 653-658
36 Smith JH. Prediction of the risk of breast cancer using computer vision techniques. PhD thesis. Department of Medical Biophysics, University of Manchester, UK, 1998
37 Marchant TE. Calibrating densities in mammography MSc Thesis. Department of Imaging Science and Biomedical Engineering, University of Manchester, 2002
42 Kopans DB, Smith RA, Duffy SW. Mammographic Screening and “Overdiagnosis” Radiology 2011; 260: 1-5
43 Kopans DB. The 2009 U.S. Preventive Services Task Force guidelines ignore important scientific evidence and should be revised or withdrawn. Radiology 2010; 256:15–20


Heggie JCP, Liddell NA, Maher KP. Applied Imaging Technology, St Vincents Hospital, Melbourne, 4th edition 2001


Accessed 2 April 2012
77 Ursin G, Parisky YR, Pike MC, Spicer DV. Mammographic density changes during the menstrual cycle. Cancer Epidemiol Biomarkers Prev 2001; 10: 141-142


84 Santen R. Menopausal hormone therapies: Their effect on mammographic density and breast cancer risk. Gynecological Endocrinol 2005; 21 (Supplement 1): 12-16


92 Laloo F, Varley J, Ellis D, O’Dair L, Pharoah P, Evans DGR and the early onset breast cancer study group. Family history is predictive of pathogenic mutations in BRCA1, BRCA2 and TP53 with high penetrance in a population based study of very early onset breast cancer. Lancet 2003; 361:1011-1012
98 Lee AHS. Why is carcinoma of the breast more frequent in the upper outer quadrant? A case series based on needle core biopsy diagnoses. The Breast 2005; 14: 151-152
100 Darbre PD. Recorded Quadrant Incidence of Female Breast cancer in Great Britain suggests a disproportionate increase in the Upper Outer Quadrant of the breast. Anticancer research 2005; 25: 2542-2550
108 Butler LM, Gold EB, Conroy SM, Crandall CJ, Greendale GA, Oestreicher N, Quesenberry Jr CP, Habel LA. Active, but not passive cigarette smoking was inversely associated with mammographic density. Cancer Causes Control 2010; 21:301-311


DICOM website. Available at medical.nema.org

Accessed 17 September 2012


Harvey JA. Quantitative assessment of percent breast density: analog versus digital acquisition. Technology in Cancer Research and Treatment 2004; 3(6): 611-616


Patel HG. Quantifying changes in glandular tissue from mammograms. MSc Thesis. Department of Computer Science, University of Manchester, 2003

IPEM Report 89. Commissioning and Routine Testing of Mammographic X-ray Systems. Institute of Physics and Engineering in Medicine, 2005


Accessed 22 November 2005


Constantinou C, Attix FH, Paliwal BR. A solid water phantom for radiotherapy x-ray and \(\gamma\)-ray beam calibrations. Medical Physics 1982; 9: 436

Hammerstein GR, Miller DW, White DR, Masterson ME, Woodard HQ, Laughlin JS. Absorbed radiation dose in mammography. Radiology 1979; 130: 485 - 491


Accessed 17 March 2008


181 Ashbrook A, Thacker NA. Tutorial: Algorithms For 2-Dimensional Object Recognition. Last updated 1/12/1998. Lecture notes from Computer Vision module, Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester.


Personal correspondence with Claire Mercer, Superintendent Radiographer at Bolton Breast Screening Centre, 19 March 2011.


GeoConvert is undertaken at the Census Dissemination Unit, Mimas, University of Manchester as part of the Economic and Social Research Council’s Census Programme.


207 Health Survey for England 2007: Latest trends. Available at
Accessed 20 February 2012
209 Abridged Clinical Results, Quantra Volumetric Density Software. Available at
210 NHS Cancer Reform Strategy, Department of Health December 2007. Available at
Accessed 14 September 2012
212 NHSBSP Technical Evaluation reports. Available at
## Appendix: Documentation from the Feasibility Study

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Participant Information Sheet</td>
<td>248</td>
</tr>
<tr>
<td>A2</td>
<td>Study Invitation</td>
<td>250</td>
</tr>
<tr>
<td>A3</td>
<td>Questionnaire 1</td>
<td>251</td>
</tr>
<tr>
<td>A4</td>
<td>Questionnaire 2</td>
<td>252</td>
</tr>
</tbody>
</table>
Breast Density Measurement from Mammograms

PARTICIPANT INFORMATION SHEET

The purpose of the ‘Breast Density’ study is to evaluate, in the breast screening programme, the practicality and accuracy of a method for measuring the amount of dense tissue within a breast. Breast density has been shown to be strongly related to breast cancer risk.

If you participate in the Breast Density study:

- Your mammogram will be taken as normal when you attend for breast screening
- No additional visits are required (unless an abnormality is detected)
- Your mammogram will be read by two specialists in the normal way. The timescale of your diagnosis will not be affected in any way
- Research staff at the University of Manchester and the Christie Hospital will calculate the amount of dense tissue from your mammogram and analyse the data on your questionnaire

This study has been given a favourable ethical opinion by the Bolton Local Research Ethics Committee.

Currently, invitation for breast screening is based purely on age or a family history of the disease. These are known as ‘risk factors’. Previous clinical studies have shown a strong link between the amount of dense (non-fatty) tissue within the breast and the development of breast cancer, suggesting that breast density may also be a significant risk factor.

A method for measuring the amount of dense tissue within a breast, based on the x-ray image (the mammogram), has recently been developed at the University of Manchester and the Christie Hospital.

We are inviting women who attend the NHS breast screening programme (NHSBSP) to take part in a study to evaluate how acceptable the method is, both in terms of its practicality and its accuracy. We plan to recruit 3,000 women.

What is the purpose of the study?
A special property which sets breast density apart from other risk factors is that it can be altered. Diet, exercise, and hormone therapy have all been shown to reduce breast density. It is possible that a reduction in breast density corresponds to a reduction in breast cancer risk, although this has yet to be proved. In order to do this, we must find a method which is suitable for use in the screening programme and capable of accurately measuring the amount of dense tissue within the breast. The purpose of this study is to test a new method to see if it is suitable.

Do I have to take part?
No. It is up to you to decide whether or not to take part. Before you decide, please read this information leaflet and ask us if there is anything that is not clear or if you would like more information. If you decide not to take part, you do not need to give a reason, and the standard of care you receive will not be affected.
How do I volunteer to take part or withdraw from the study?
If you would like to take part, you should complete the enclosed questionnaire. Please take your completed questionnaire with you when you attend your screening appointment and inform the radiographer at the start of the examination. She will ask you to sign a ‘consent form’.
If you subsequently decide that you do not wish your mammograms to be included in the study, please contact the research staff using the details given below.

What will happen to me if I take part?
There will be no difference in your visit to breast screening and your mammograms will be taken and looked at as normal. No additional visits are required (unless an abnormality is detected on your mammogram). In order to calculate the breast density it is necessary to place an additional marker near the breast so it appears in the mammogram and some very small metal discs on top of the compression device. These objects will not come into contact with your breasts. They will appear on the mammogram but it is extremely unlikely that they will obscure any important details.

How will my treatment be affected?
Your treatment will not be affected in any way. The specialists reading your films will not be carrying out any density analysis so diagnosis will not be delayed. All density analysis will be carried out at a later date by research staff at the University and the Christie Hospital in Manchester.

What are the possible disadvantages of taking part?
There is a possibility that on the mammograms of women with larger breasts, the images of the metal discs may overlap the image of the edge of the breast by a very small amount. Abnormalities only rarely occur in this area, and the ability of the specialist to make a diagnosis will not be significantly affected. The discs have a diameter of only 3mm. A previous study showed that the discs overlapped a small amount of breast tissue in only 2.5% of cases. Important details were not missed as even on the rare occasion when both views were affected, a marker would not obscure the same area of tissue on both views.

What are the possible benefits of taking part?
Whilst there are no direct benefits for the individuals involved in this research, the project could benefit women in the future. If successful, the method could be used to identify those women at higher risk of breast cancer. These women could be invited for more frequent mammograms or be informed about methods for reducing their breast density. This could mean a potential reduction in the risk of getting cancer, or earlier detection, both of which would offer improved chances of survival.

What if something goes wrong?
If you wish to complain or have any concerns about any aspect of the way you have been approached or treated, the normal NHS complaints mechanisms would be available to you.

Will my information be kept confidential?
Yes, any information collected about you during this study will be kept strictly confidential. We will identify any information about you by a study number known only to the research team.

What will happen to the results of the study?
The results of the study will be published in medical journals. To do this we have to ensure that our study information is double-checked by other professionals in research and healthcare. If you would like a general summary of the results of the study (due 2008) please contact the research team or your local breast screening centre.

For further information on any aspect of the study contact:
Jenny Diffey (Clinical Scientist)
North Western Medical Physics
Christie Hospital NHS Trust
Manchester Tel: 0161 4463537 Mobile: 07925 125874
M20 4BX E-mail: jenny.diffey@physics.cr.man.ac.uk

Thank you, we hope you will agree to take part in this study
Dear Madam,

You were recently invited for a routine screening mammogram.

I am writing to let you know that your screening centre is currently taking part in a research study to evaluate a method of measuring the amount of dense tissue within a breast. Breast density has been shown to be strongly linked to the risk of breast cancer. We would like to give you the opportunity to have your screening mammogram included in this study.

Information about this research is included. If you would like to take part, all you need to do is fill in the questionnaire on the back of this letter and take it with you when you attend your screening appointment, where you will be asked to sign a consent form.

There will be no difference in your visit for breast screening and your results will be sent to you as normal.

If you have any questions about the study please call the research team before your appointment on telephone number 0161 4463537 or 07925 125874.

Thank you for taking the time to read this information.

Yours faithfully,

Jenny Diffey
Clinical Scientist

Research funded by:
Participant Questionnaire 1

Please complete all sections and tick the box which represents how certain you are about each fact.

Name: ………………………………………………………………………

Date of Birth: ……………………………………………………………

How sure are you?

Height: .......... ft .......... in   OR       .......... cm

- Don’t know
- Not sure
- Quite sure
- Certain

Weight: .......... st .......... lb   OR       .......... kg

- Don’t know
- Not sure
- Quite sure
- Certain

Age at first period: ........................

- Don’t know
- Not sure
- Quite sure
- Certain

Age at menopause: ........................

- Don’t know
- Not sure
- Quite sure
- Certain

Age when first child born: ........................

- Don’t know
- Not sure
- Quite sure
- Certain

Has your mother or a sister had breast cancer? ........................................

- Don’t know
- Not sure
- Quite sure
- Certain

If so, what age were they? ........................

- Don’t know
- Not sure
- Quite sure
- Certain

Ethnicity (Please tick one box)

<table>
<thead>
<tr>
<th>Asian or Asian British – Bangladeshi</th>
<th>Mixed – White &amp; Black African</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian or Asian British – Indian</td>
<td>Mixed – White &amp; Asian</td>
</tr>
<tr>
<td>Asian or Asian British – Pakistani</td>
<td>Mixed – White &amp; Black Caribbean</td>
</tr>
<tr>
<td>Other Asian background</td>
<td>Other Mixed background</td>
</tr>
<tr>
<td>Black or Black British – African</td>
<td>White - British</td>
</tr>
<tr>
<td>Black or Black British – Caribbean</td>
<td>White - Irish</td>
</tr>
<tr>
<td>Other Black background</td>
<td>Other White background</td>
</tr>
<tr>
<td>Chinese</td>
<td>Other Ethnic background Please specify:</td>
</tr>
<tr>
<td></td>
<td>............................................................</td>
</tr>
</tbody>
</table>
Participant Questionnaire 2

Please try and give an accurate answer to each question. If you don’t know, just leave it blank. If you can’t remember dates or other information precisely, please do the best you can and tick the box that shows how certain you are about your answer.

Please return your completed questionnaire in the enclosed pre-paid envelope. Thank you

<table>
<thead>
<tr>
<th>Name:</th>
<th>Postcode:</th>
<th>Date of birth:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Height:</th>
<th>Weight:</th>
<th>How sure are you?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ft/in</td>
<td>lb/kg</td>
<td>Not sure □</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fairly sure □</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Certain □</td>
</tr>
</tbody>
</table>

At what age did you have your first period? □

Have you been through the menopause yet? Yes □ No □

If not, are you currently going through the menopause? Yes □ No □

If you answered ‘yes’ to either of the above, at what age did you start going through the menopause? □

Have you ever been on Hormone Replacement Therapy (HRT)? Yes □ No □

If yes, for how many years? □

Are you still on HRT? Yes □ No □

If not, how long ago did you stop? □

Have you ever used natural HRT remedies (e.g. red clover, black cohosh, natural progesterone, wild yam)? Yes □ No □

If yes, which product? □

and for how many years? □

Are you still using it? Yes □ No □

If not, how long ago did you stop? □

PLEASE TURN OVER
<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
<th>Not sure</th>
<th>Fairly sure</th>
<th>Certain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you ever smoked?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If yes, how many per week?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and for how many years?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you still smoke?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If not, how long ago did you stop?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have you had any children?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If yes, how many?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How old were you at your first pregnancy?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has your mother or a sister had breast cancer?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If yes, please specify whether mother and/or sister(s) and state what age they were</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Ethnicity (Please tick one box)                                          |     |    |          |             |         |
| Asian or Asian British - Bangladeshi                                     |     |    |          |             |         |
| Asian or Asian British - Indian                                          |     |    |          |             |         |
| Asian or Asian British - Pakistani                                       |     |    |          |             |         |
| Other Asian background                                                   |     |    |          |             |         |
| Black or Black British - African                                         |     |    |          |             |         |
| Black or Black British - Caribbean                                       |     |    |          |             |         |
| Other Black background                                                   |     |    |          |             |         |
| Chinese                                                                  |     |    |          |             |         |
| Mixed - White and Asian                                                 |     |    |          |             |         |
| Mixed - White and Black African                                         |     |    |          |             |         |
| Mixed - White and Black Caribbean                                       |     |    |          |             |         |
| Other Mixed background                                                  |     |    |          |             |         |
| White - British                                                         |     |    |          |             |         |
| White - Irish                                                           |     |    |          |             |         |
| Other White background                                                  |     |    |          |             |         |
| Other Ethnic background - please specify                                 |     |    |          |             |         |

- 253 -