



University of
Salford
MANCHESTER

Impact of errors in recorded compressed breast thickness measurements on volumetric density classification using volpara v1.5.0 software

Waade, G, Highnam, R, Hauge, I, McEntee, M, Hofvind, S, Denton, E, Kelly, J, Sarwar, J and Hogg, P

<http://dx.doi.org/10.1118/1.4948503>

Title	Impact of errors in recorded compressed breast thickness measurements on volumetric density classification using volpara v1.5.0 software
Authors	Waade, G, Highnam, R, Hauge, I, McEntee, M, Hofvind, S, Denton, E, Kelly, J, Sarwar, J and Hogg, P
Type	Article
URL	This version is available at: http://usir.salford.ac.uk/id/eprint/38834/
Published Date	2016

USIR is a digital collection of the research output of the University of Salford. Where copyright permits, full text material held in the repository is made freely available online and can be read, downloaded and copied for non-commercial private study or research purposes. Please check the manuscript for any further copyright restrictions.

For more information, including our policy and submission procedure, please contact the Repository Team at: usir@salford.ac.uk.

Error in recorded compressed breast thickness measurement impacts on volumetric density classification using Volpara v1.5.0 software

Gunvor Gipling Waade

Oslo and Akershus University College of Applied Sciences, Oslo, Norway

5 *School of Health Sciences, University of Salford, Salford, UK*

Ralph Highnam

Matakina Technology Limited, Wellington, New Zealand

10 Ingrid H. R. Hauge

*The Intervention Centre, Rikshospitalet, Oslo University Hospital, Oslo, Norway*⁵*Discipline of Medical*

Mark F. McEntee

15 *Discipline of Medical Radiation Science, Faculty of Health Sciences, University of Sydney, Australia*

Solveig Hofvind

Cancer Registry of Norway, Oslo, Norway

Oslo and Akershus University College of Applied Sciences, Oslo, Norway

20

Erika Denton

Norfolk and Norwich University Hospital NHS Foundation Trust, Norfolk, UK

Judith Kelly

25 *Countess of Chester Hospital NHS Foundation Trust, Chester, UK*

Jasmine J. Sarwar

School of Health Sciences, University of Salford, Salford, UK

30 Peter Hogg

School of Health Sciences, University of Salford, Salford, UK
Karolinska Institute, Stockholm, Sweden

Abstract

35 **Purpose:** Mammographic density has been demonstrated to predict breast cancer risk. It has been proposed that it could be used for stratifying screening pathways and recommending additional imaging. Volumetric density tools use the recorded compressed breast thickness (CBT) of the breast measured at the x-ray unit in their calculation, however the accuracy of the recorded thickness can vary. The aim of this study was to investigate whether inaccuracies in recorded CBT impact upon volumetric density classification and to examine whether the current quality control (QC) standard is sufficient for assessing mammographic density.

40 **Methods:** Raw data from 52 digital screening mammograms were included in the study. For each image, the clinically recorded CBT was artificially increased and decreased to simulate measurement error. Increments of 1mm were used up to $\pm 15\%$ error of recorded CBT was achieved. New images were created for each 1mm step in thickness resulting in a total of 974 images which then had Volpara Density Grade (VDG) and volumetric density percentage assigned.

45 **Results:** A change in VDG was recorded in 38.5% (n= 20) of mammograms when applying $\pm 15\%$ error to the recorded CBT and 11.5 % (n= 6) were within the QC standard prescribed error of $\pm 5\text{mm}$.

50 **Conclusion:** The current QC standard of $\pm 5\text{mm}$ error in recorded CBT creates the potential for error in mammographic density measurement. This may lead to inaccurate classification of mammographic density. The current QC standard for assessing mammographic density should be reconsidered.

55 Key words: paddle, error, mammogram, QC

I. INTRODUCTION

Mammographic density (MD) is the radiographic density of the breast on a mammogram determined by the composition of breast tissue, the amount of radiodense (parenchymal and connective tissue) and radiolucent tissue (fat).^{1,2} MD is a strong risk factor for breast cancer, where the risk of developing breast cancer is three to six times greater for women with extremely high density compared to those with fatty breast.¹⁻⁵ Measurement of MD is important for breast cancer risk prediction and might be used for imaging pathway or screening interval recommendations. The masking effect from increased MD also reduces the mammography screening sensitivity.³ Women with high MD might benefit from additional imaging, such as ultrasound (US) or magnetic resonance imaging (MRI), or more frequent screening compared with women with low MD.⁴⁻⁷

There are several methods to measure MD⁸⁻¹³, the most common of which is visual assessment by radiologists using the Breast Imaging Reporting and Data System (BIRADS) scale.⁸ This system is prone to inter and intra-reader variability^{14, 15} and thus several computer-assisted methods have been developed including Cumulus⁹ and Madena¹⁰. These methods all rely on human interpreters/readers to set the threshold for dense tissue. However, these methods have shown reduced subjectivity compared to the visual techniques.¹⁶ Automation of this process removes human variability. Several automated systems exist, including AutoDensity¹¹, Quantra¹² and Volpara¹³. The latter two use volumetric breast density assessment and has been cleared by the Food and Drug Administrative (FDA) as adjunctive supporting tools.¹⁷ VolparaDensity software (*Matakina Technology Limited., Wellington, New Zealand*)¹³ uses the raw image and meta data from digital mammograms to calculate average volumetric breast density percentage (VBD%).^{13,18,19} Volpara estimates VBD% by dividing the volume of fibroglandular tissue by the total volume of the breast, as follows;

$$VDB\% = \frac{\text{Volume of fibroglandular tissue}}{\text{Total volume of the breast (area of the breast x recorded breast thickness)}}$$

The total volume of the breast is found by multiplying the area of the breast with the recorded thickness of the compressed breast, with correction for the uncompressed breast edge region. The recorded CBT is used to calculate the volume of fibroglandular tissue, but much more explicitly used in the calculation of breast volume. The accuracy of CBT specified by manufacturers ranges between $\pm 5-10$ mm²⁰, and any error inherent in the measurement will result in inaccuracies in VBD%. Errors in recorded CBT from mammography machines are expected to be the largest contributing factor for the MD algorithm's inaccuracies.¹⁶

A typical quality control (QC) tolerance level is up to ± 5 mm difference between recorded and measured CBT.²¹ In a previous study our group investigated the accuracy of recorded CBT for a range of screen film mammography and full-field digital mammography (FFDM) units using a deformable phantom.²⁰ The recorded CBT varied up to 14.3% (5.6mm) and 26.4% (10.5mm) from measured thickness for non-flexible and flexible (rigid) paddles respectively when applying 100 Newton (N) compression force.²⁰ We noted that techniques exist to detect and correct for compression plate slant.²² However, the described error in recorded CBT may lead to inaccurate estimates of VBD%. This could lead to incorrect classification of women into specific MD groups, and being assigned to an incorrect imaging pathway or screening interval. Accurate CBT measurement in mammography is also important in order to calculate mean glandular dose (MGD).^{23,24} The aim of this study was to investigate the impact of errors

105 in recorded CBT on VBD classification and examine whether the current QC standard is sufficient for assessing mammographic density.

II. METHODS AND MATERIAL

The study was approved as Service Evaluation at the Countess of Chester Hospital, Chester, UK (Reference number: ID 3763). Raw data from 52 digital breast screening mammograms (Hologic Selenia Dimensions Mammography machine) were included in the study. For each image, the simulated thickness was increased and decreased in increments of 1mm until $\pm 15\%$ from the recorded CBT was reached. 15% change in recorded CBT was used as this was the previously reported error in recorded CBT for non-flexible paddles.²⁰ New images were created for each mm step in CBT resulting in a total of 974 images. All images were then assessed for MD using Volpara v1.5.0 software (Matakina International Ltd, Wellington, NZ).

II.A. Selection of images:

The 52 mammograms used in the study were chosen from a selection of 300 mammograms from 300 women aged 50-69 years participating in the NHS Breast Screening Programme (NHSBSP) during 2014. One mammogram of each woman was available and there was a mix of left/right breast and cranial-caudal (CC)/ mediolateral oblique (MLO) views (n=26 CC-images and 26 MLO-images). Two experienced mammography image readers reviewed the images individually and independently on 5 megapixel (MP) monitors (Hologic SecureViewDX Diagnostic Workstation) under standard reporting conditions for technical quality, positioning, artefacts, pathology and blur. Blur was assessed by confirming that breast anatomical structures had distinct/sharp edges.^{25,26} None of the images included known pathology at time of reading. Images also had passed routine clinical processes for technical quality of breast screening within the UK.²⁷ A total of 100 images met the inclusion criteria. 52 of these were chosen based on a consensus meeting between the readers as representing an equal distribution of breast sizes and the BI-RADS density classification grades.⁸

II.B. Changing the recorded CBT in the Digital Imaging and Communications in Medicine (DICOM) header

The recorded CBT in the DICOM file header was used as a baseline from which the thickness was adjusted by $\pm 15\%$ in 1mm steps. The thickness was adjusted using a software known as DVTk DICOM Editor Tool 3.2.6.²⁸ The adjusted CBT was rounded off to the closest whole mm. This created between 10 and 24 new image datasets for each original image resulting in 974 images. The thickness of the baseline compressed breast determined how many images that were created for each image; thicker breasts resulted in more images.

II.C. Mammographic density classification

Raw data from the 974 image datasets was then processed using VolparaDensity v1.5.0 software to estimate average MD percentages (VBD%) and Volpara Density Grades (VDG). VDG is a BIRADS 4th Edition Density Category, and is obtained by simply thresholding the average volumetric density for the study: VDG 1 = 0-4.5% VBD%, VDG2 = 4.5-7.5% VBD%, VDG3 = 7.5-15.5% VBD%, and VDG4 >15.5% VBD%.²⁹ The range and mean of change in VBD% were calculated for all mammograms. To investigate how percentage error in recorded CBT affected the VBD%, the changes in thickness (mm) were calculated as a percentage; a 1mm step on a 60mm breast was calculated as a 2% error in recorded CBT ($1\text{mm}/60\text{mm} \times 100 = 1.667 = 2\%$). The maximum error in recorded CBT was 15% in all cases.

II.D. Statistics

155 Statistical analyses were conducted using the statistical package R (R for Mac OS, version 3.0.2 GUI 1.62 Snow Leopard build (6558)).³⁰ Data were grouped by patient, recorded CBT, change in recorded CBT (\pm mm), total breast volume, total volume of fibroglandular tissue, VBD% and VDG. To investigate how the estimated density varied with the recorded breast thickness, breast volume and fibroglandular volume, density (D) was modelled as:

1. a function of CBT (t)
2. a function of CBT and breast volume (V_B)
- 160 3. a function of CBT and fibroglandular volume (V_f)
4. a function of CBT, breast volume and fibroglandular volume

III. RESULTS

165 Total breast volume and volume of fibroglandular tissue increased while VBD% decreased
with increasing recorded thickness. The correlation r^2 between image parameters and
estimated MD was strongest for estimated MD as a function of CBT, breast volume and
fibroglandular volume (0.81).

170 The correlation was 0.68, 0.33 and 0.30 for MD as a function of CBT and fibroglandular
volume, MD as a function of CBT and breast volume and MD as a function of CBT,
respectively.

III.A. Changes in VBD% when applying error to the recorded CBT breast thickness

175 The changes in VBD% were greater when decreasing compared to increasing the recorded
CBT. There was no difference in mean change in VBD% by CC or MLO view; the values
were the same as for the views combined. The largest change in VBD% was 2.5 and 3.1 for
5mm and 15% respectively when decreasing the recorded CBT (Table 1).

Table 1: The mean and largest changes in Volumetric breast density (VBD%) when applying
5mm and 15% error in recorded compressed breast thickness (CBT).

Changes in VBD% by error in compressed breast thickness (CBT)	5 mm	15 %
Mean change	0.5	0.8
Mean change when decreasing	0.5	0.9
Largest change when decreasing	2.5	3.1
Mean change when increasing	-0.4	-0.7
Largest change when increasing	-2.0	-2.3

180

III.B. Volpara density grade

185 There were 13 mammograms classified with VDG 1, 22 with VDG 2, 14 with VDG 3, and 3
with VDG 4. The changes in VBD% increased with increasing original VDG (Figure 1).
Mean change in VBD was 0.2 and 0.4 at 5mm and 15% error in compressed breast thickness
for VDG1 while it was 2.0 and 2.3 at 5mm and 15% error for VDG4 (Table 2).

Changes in volumetric breast density (VBD%) by change in thickness for the different Volpara density grades (VDG)

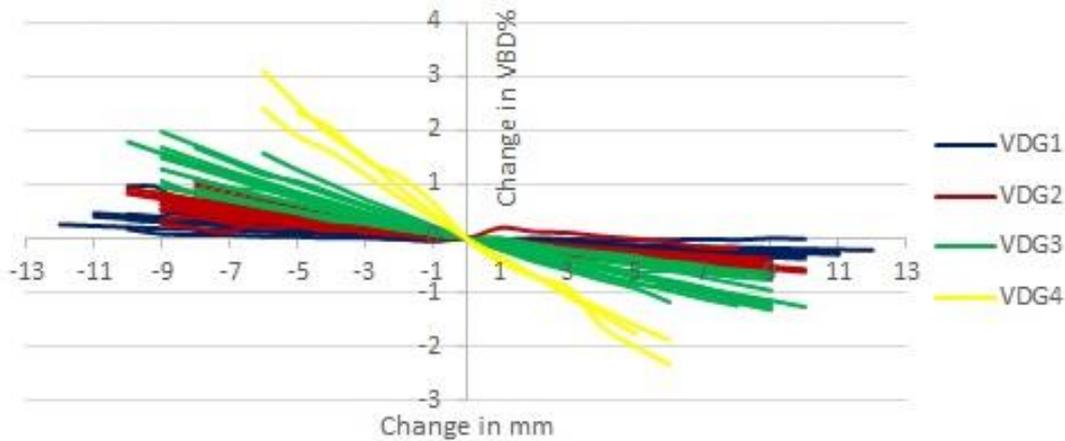


Figure 1: Changes in volumetric breast density (VBD%) for each original image by Volpara density grade (VDG1-4) when increasing and decreasing the recorded compressed breast thickness (CBT) by 1mm up to 15% error from the recorded CBT.

Table 2: The mean and largest changes in Volumetric breast density (VBD%) when applying 5mm and 15% error in recorded compressed breast thickness (CBT) for the different Volpara density grades (VDG1-4).

Changes in VBD% by error in compressed breast thickness (CBT)	VDG1		VDG2		VDG3		VDG4	
	5 mm	15 %						
Mean change	0.2	0.4	0.3	0.6	0.7	1.2	2.0	2.3
Largest change when decreasing	0.3	1.0	0.6	1.0	1.3	2.0	2.5	3.1
Largest change when increasing	-0.2	-0.4	-0.5	-0.7	-0.9	-2.0	-2.0	-2.3

In total 20 out of 52 mammograms changed their density grade when a 15% error was applied to the recorded CBT (Table 3). Fifteen mammograms increased one density group when reducing the CBT 15% and five decreased one density group when increasing the CBT 15%. Most changes were from VDG 2 to 3 (n=8). Six changes in VDG occurred within the 5mm QC guidelines. For the different projections (CC/MLO), twelve changes in VDG were in CC and eight changes were in MLO.

205 **Table 3:** Changes in Volpara Density Grade (VDG) when applying 15% and 5mm error in
 recorded CBT. Left column shows changes in VDG. The two following columns show
 number of mammograms with changes in VDG and average change in VBD% and standard
 deviation for 5mm change in recorded CBT. The two columns on the right includes the
 number of mammograms with changes in VDG with average change in VBD% and standard
 210 deviation for 15% change in recorded CBT.

Change in VDG	5mm error in compressed breast thickness (CBT)		15% error in compressed breast thickness (CBT)	
	(n=6)	Average change in VBD% (SD)	(n=20)	Average change in VBD% (SD)
from 1 to 2	1	0.3 (N/A)	4	0.5 (0.27)
from 2 to 1	-	-	2	0.3 (0.01)
from 2 to 3	2	0.4 (0.28)	8	0.6 (0.23)
from 3 to 2	1	0.3 (N/A)	3	0.5 (0.19)
from 3 to 4	2	0.8 (0.2)	3	0.9 (0.17)
from 4 to 3	-	-	-	-

IV. DISCUSSION

Our study identified that error in CBT had an impact on the estimated volumetric breast density. 38.5% of the mammograms had a change in VDG when applying $\pm 15\%$ error to the recorded CBT and 11.5 % (n= 6) had a change in VDG within the QC standard prescribed error of $\pm 5\text{mm}$. There were larger changes in VBD% for the mammograms with the highest original VDG.

We identified a high correlation between MD and a combination of CBT, breast volume and fibroglandular volume ($r^2=0.81$). The image parameters (CBT, breast volume and fibroglandular volume) had limited effect on the estimated density alone. The findings are as expected, as the software estimates VBD% by dividing the volume of fibroglandular tissue by the total volume of the breast, found by multiplying the area of the breast by the recorded compressed breast thickness.

QC guidelines indicate all under- and over-estimated measures of CBT outside $\pm 5\text{mm}$ of the recorded CBT are considered equal faults. However, our data indicated that underestimation of CBT has a greater impact on the estimated MD than overestimation (mean change in VBD was 0.5 and 0.9 for 5mm and 15% when decreasing the CBT compared to -0.4 and 0.7 when increasing). This may be due to the total volume of the breast being directly affected by the CBT whereas the volume of fibroglandular tissue is not as affected by the change in CBT.

In 2005, Blot & Zwiggelaar³¹ demonstrated that estimations of MD using the h_{int} model relied heavily on accurate measurements, in particular recorded CBT. They state that the CBT must be estimated within 0.5mm to obtain an error in MD smaller than 5%. Tyson *et al.*³² stated that a mean accuracy of less than 1 mm is required to make good estimates for the VBD. In our study, the largest change in VBD% was 2.5 for 5mm and 3.1 for 15% error related to recorded CBT. This might indicate that newer volumetric methods for estimating density are more robust than older.

Where errors in CBT occur in a clinical setting, the technical factors to form the mammogram (tube voltage [kVp], target material and filter material) might change due to the use of automatic exposure control (AEC)³³, and this in turn would affect the VBD³⁴. Feng et al.³³ found the tube voltage to change with approximately 1kVp per cm in CBT and a shift from Rhodium to Silver filter from 6 cm to 7 cm CBT.³³ If the CBT is lower than actual thickness, the AEC would choose a lower kVp, and thus the estimated VBD would be overestimated (increased volume of fibroglandular tissue due to lower kVp in the numerator and decreased volume of breast tissue due to decreased thickness in denominator). In contrast, if the CBT is higher than actual thickness, the AEC would choose a higher kVp, and thus the estimated VBD would be underestimated. Lau et al³⁴ investigated how errors in the recorded imaging physics parameters affected the VBD, by changing the recorded CBT, kVp, exposure (mAs), target material, filter material and filter thickness, in addition to simulating changes in detector gain and offset by adjusting pixel values.³⁴ They found the exposure, detector gain and filter thickness to have a negligible or no impact on the VBD, while simulated errors in tube voltage, target material, filter material, detector offset and compressed breast thickness had a significant impact on the VBD.³⁴ From this point it is safe to anticipate that in a clinical setting the AEC would affect our results.

260 The use of MD information has not yet been standardized. The ACRIN 6666 trial confirms
the utility of using ultrasound³⁵ and MRI³⁵ for women with densest breasts, and additional
imaging for women with high MD and higher risk of developing cancer has been
suggested.^{36,37} Women with lower familial risk, no genetic markers and lower mammographic
density could benefit from less frequent imaging and no additional imaging outside the
265 standard mammographic screening program.^{38,39} Additional imaging and/or less/more
frequent screening obviously is associated with higher costs and potential for increase false-
positive results, and future screening regimes need to consider this when considering stratified
screening programs.^{35,36,40}

270 Although changes in VBD% are important, they become more so with clinical use of MD
groups. All changes in VDG of our study occurred within 10mm (15%) error in recorded
CBT. A ± 5 mm error in recorded CBT is considered acceptable in clinical practice as ± 5 mm
is the QC standard in UK²¹; this level of inaccuracy in VBD% is likely to already be
occurring. It is possible that the problem is worse than that, as CBT accuracy specified by
275 manufacturers ranges between ± 5 -10 mm²⁰. With MD category systems, it is important to
place a woman in the appropriate density group if the groups are assigned different imaging
pathways. Previous studies comparing annually and biennially breast cancer screening
intervals found higher probability of false-positive recalls and/or biopsy recommendations for
women being screened annually to biennially.^{41,42} If a woman is misplaced from low density
280 to a high density group, this might lead to anxiety for the woman, possibly increased cases of
false positive results and increased possibility of unnecessary tests and potentially
overtreatment. This might also lead to false sense of security and loss of confidence in the
screening program. Equally, if she moves from high density to low density group then a
cancer might be missed by not performing supplemental screening.

285 In our study, changing the read out thickness had little impact on the VBD% for VDG1 and
VDG2. The change in VBD% was highest for VDG4, which means that these women are at
highest risk of receiving an inaccurate MD group assignment. Mean change in density for all
mammograms was 0.8 VBD% at $\pm 15\%$ error which might lead to a change in density group if
the VBD% is close to the boundary of a VDG. This number was reduced to 0.5 VBD% when
290 having ± 5 mm (the QC standard) as the maximum error. The use of density categories rather
than a continuous scale means that women who are placed in between two groups could
move, relatively easily, between categories when errors in recorded CBT occur. It might be
better to use a continuous measure of MD in the future than using a category approach, both
because risk is continuous, but also because it removes the error from arbitrary categorization.

295 Our study was solely based on the results of the Volpara software. Both Volpara^{13,18,19} and
Quantra^{12,43} calculates VBD% by comparing each pixel's attenuation to the attenuation of
pixels that are labeled as entirely fatty tissue (pixels with the lowest attenuation), and then
divides the volume of fibroglandular tissue by the total breast volume⁴⁴. However, the
300 systems have some differences such as internal calibration¹⁴ and correction for compression
paddle height and tilt⁴⁵. Studies have reported moderate to excellent correlations (Pearson's
correlation coefficient [r^2] = 0.78-0.99, Intra class correlation [ICC] = 0.64-0.96) of Volpara
and Quantra.⁴⁵⁻⁴⁷ Quantra has shown higher values of total volume of fibroglandular tissue
and VBD^{44,45}, while Volpara has shown higher total breast volume^{44,47}. As the automated
305 systems have different algorithms, the outcome would probably differ. However, we assume
that all density assessment algorithms including the CBT is affected by incorrect thickness
information; although the size of the effect is difficult to predict.

310 Previous work by our group has shown that the recorded CBT was different to the actual CBT
and varied by up to 26.4% from measured CBT for flexible paddles.²⁰ Although Volpara and
other density measurement systems determine and correct for tilt, this level of variation in
measured CBT will result in variation in VBD% and can result in misclassification of MD
groups. In this study, a 15% change in CBT resulted in a change in CBT between 5 mm and
315 12mm, showing that the current QC standard of 5mm might not be complied in clinical
practice. This further raises questions whether the QC standard of ± 5 mm can be achieved by
flexible paddles, which again raises important questions about accuracy when assessing MD
with equipment with flexible paddles. There is an increasing frequency of clinical questions
around flexible paddles^{22,48} and it would appear that moving forward, the QC around paddles
has to be improved in order to obtain higher quality MD measurements. As the CBT is of
320 importance both for MD estimations, and for the estimation of MGD, there is a lot to gain in
the accuracy of these estimates by tightening the QC standard. Currently, flexible paddles
should only be used with the caveat that this can result in an inaccurate measurement of MD.

325 Our study included 52 mammograms. Sample size estimation demonstrated that in order to
detect an effect size of 8mm with a power of 0.8, 503 mammograms would be needed. A
further limitation to the study is that we based the estimation of error to the recorded CBT by
changing the recorded thickness in the DICOM header only. In a clinical situation, differences
in compression on the same breast are likely to have a greater effect on the height and area of
the breast. These effects are operator dependent and more difficult to control. Further studies
330 might include increasing and decreasing the kVp to see the effect on MD, repeating the study
with a larger data set and include a larger interval of errors up to 25% to include error from
flexible paddles.

V. CONCLUSION

335 Variations in recorded CBT impact upon the accuracy of MD estimations. As flexible paddles can have variations of 25% in CBT, these paddles should be used with care when subsequent assessment of MD is likely. The current QC standard of ± 5 mm error in recorded CBT creates the potential for error in mammographic density measurement. This may lead to inaccurate classification of mammographic density. The current QC standard of ± 5 mm for assessing mammographic density should be reconsidered.

340 **References:**

- ¹Ursin, G., Ma, H., Wu, A. H., Bernstein, L., Salane, M., Parisky, Y. R., ... & Pike, M. C. (2003). Mammographic density and breast cancer in three ethnic groups. *Cancer Epidemiology Biomarkers & Prevention*, *12*(4), 332-338.
- 345 ²Boyd, N. F., Martin, L. J., Bronskill, M., Yaffe, M. J., Duric, N., & Minkin, S. (2010). Breast tissue composition and susceptibility to breast cancer. *Journal of the National Cancer Institute*, *102*(16), 1224-1237.
- ³Boyd, N. F., Guo, H., Martin, L. J., Sun, L., Stone, J., Fishell, E., ... & Yaffe, M. J. (2007). Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine*, *356*(3), 227-236.
- 350 ⁴Boyd, N. F., Byng, J. W., Jong, R. A., Fishell, E. K., Little, L. E., Miller, A. B., ... & Yaffe, M. J. (1995). Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *Journal of the National Cancer Institute*, *87*(9), 670-675.
- ⁵McCormack, V. A., & dos Santos Silva, I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiology Biomarkers & Prevention*, *15*(6), 1159-1169.
- 355 ⁶Yaffe, M. J., Boone, J. M., Packard, N., Alonzo-Proulx, O., Huang, S. Y., Peressotti, C. L., ... & Brock, K. (2009). The myth of the 50-50 breast. *Medical physics*, *36*(12), 5437-5443.
- ⁷Gubern-Mérida, A., Kallenberg, M., Platel, B., Mann, R. M., Martí, R., & Karssemeijer, N. (2014). Volumetric breast density estimation from full-field digital mammograms: a validation study. *PloS one*, *9*(1).
- 360 ⁸D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology; 2013.
- ⁹Byng, J. W., Boyd, N. F., Fishell, E., Jong, R. A., & Yaffe, M. J. (1994). The quantitative analysis of mammographic densities. *Physics in medicine and biology*, *39*(10), 1629.
- 365 ¹⁰Palomares, M. R., Machia, J. R., Lehman, C. D., Daling, J. R., & McTiernan, A. (2006). Mammographic density correlation with Gail model breast cancer risk estimates and component risk factors. *Cancer Epidemiology Biomarkers & Prevention*, *15*(7), 1324-1330.
- ¹¹Nickson, C, et al. AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. *Breast Cancer Res.* 2013; *15*:R80.
- 370 ¹²Ciatto, S., Bernardi, D., Calabrese, M., Durando, M., Gentilini, M. A., Mariscotti, G., ... & Houssami, N. (2012). A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. *The Breast*, *21*(4), 503-506.
- ¹³Highnam, R., et al. (2010). Robust breast composition measurement-VolparaTM. In *Digital mammography: 10th International workshop, IWDM 2010, Girona, Catalonia, Spain, June 16-18 2010 – proceedings*. (pp. 342-349). Springer Berlin Heidelberg.
- 375 ¹⁴Ekpo, E. U., & McEntee, M. F. (2014). Measurement of breast density with digital breast tomosynthesis—a systematic review. *The British journal of radiology*, *87*(1043), 20140460.
- ¹⁵Ekpo, E. U., Hogg, P., Highnam, R., & McEntee, M. F. (2015). Breast composition: Measurement and clinical use. *Radiography*.
- 380 ¹⁶Alonzo-Proulx, O., Jong, R. A., & Yaffe, M. J. (2012). Volumetric breast density characteristics as determined from digital mammograms. *Physics in medicine and biology*, *57*(22), 7443.
- ¹⁷FDA U.S. Food and Drug Administration. Available: <http://www.fda.gov/>.
- ¹⁸Seo, J. M., Ko, E. S., Han, B. K., Ko, E. Y., Shin, J. H., & Hahn, S. Y. (2013). Automated volumetric breast density estimation: a comparison with visual assessment. *Clinical radiology*, *68*(7), 690-695.
- 385 ¹⁹Aitken, Z., McCormack, V. A., Highnam, R. P., Martin, L., Gunasekara, A., Melnichouk, O., ... & dos Santos Silva, I. (2010). Screen-film mammographic density and breast cancer risk: a comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. *Cancer Epidemiology Biomarkers & Prevention*, *19*(2), 418-428.
- ²⁰Hauge, I. H., Hogg, P., Szczepura, K., Connolly, P., McGill, G., & Mercer, C. (2012). The readout thickness versus the measured thickness for a range of screen film mammography and full-field digital mammography units. *Medical physics*, *39*(1), 263-271.
- 390 ²¹Moore, A.C., et al. IPEM Report 89: Commissioning and routine testing of Mammographic X-Ray systems (Institute of Physics and Engineering in Medicine, 2005).

- 395 ²²Kallenberg, M. G., van Gils, C. H., Lokate, M., den Heeten, G. J., & Karssemeijer, N. (2012). Effect of compression paddle tilt correction on volumetric breast density estimation. *Physics in medicine and biology*, 57(16), 5155.
- ²³Dance, D. R. "Monte-Carlo calculation of conversion factors for the estimation of mean glandular breast dose." *Physics in medicine and biology* 35.9 (1990): 1211.
- 400 ²⁴Highnam, R. P., Brady, J. M., & Shepstone, B. J. (1998). Estimation of compressed breast thickness during mammography. *The British journal of radiology*, 71(846), 646-653..
- ²⁵Taplin S, Rutter C, Finder C, Mandelson M, Houn F, White E. Screening Mammography: Clinical Image Quality and the Risk of Interval Breast Cancer. *AJR* 2002, 178, 797–803
- ²⁶Samei E, Performance of Digital Radiographic Detectors: Quantification and Assessment Methods. *Advances in Digital Radiography: RSNA Categorical Course in Diagnostic Radiology Physics, Radiographics*, 25, 2, 2005, pp 37–47. 9. Gale AG, Scott H. Measuring Radiology
- 405 ²⁷National quality assurance coordinating group for Radiography. Published by: NHS Cancer Screening Programmes Sheffield. National Health Service Cancer Screening Programmes (NHSBSP) Quality assurance guidelines for mammography including radiographic quality control. Report 63
- 410 (2006) ISBN 1 84463 028 5
- ²⁸DVTk DICOM Editor Tool. Available:
<http://dicom.dvtk.org/modules/wiwimod/index.php?page=Download+DICOM+Editor&cmenu=downloads>
- ²⁹Matakina Technology Limited. VolparaSolutions – Volpara Density. Available:
415 <http://volparasolutions.com/solutions/volparadensity/>.
- ³⁰R for Mac OS, version 3.0.2 GUI 1.62 Snow Leopard build (6558)) written by: Urbanek, S, Bibiko HJ, Iacus SM. Copyright © 2004-2013. The R Foundation for Statistical Computing. Available:
<https://www.r-project.org/>
- 420 ³¹Blot, L., & Zwiggelaar, R. (2005). A volumetric approach to glandularity estimation in mammography: a feasibility study. *Physics in medicine and biology*, 50(4), 695.
- ³²Tyson, A. H., Mawdsley, G. E., & Yaffe, M. J. (2009). Measurement of compressed breast thickness by optical stereoscopic photogrammetry. *Medical physics*, 36(2), 569-576.
- ³³Feng, S. S. J., & Sechopoulos, I. (2012). Clinical digital breast tomosynthesis system: dosimetric characterization. *Radiology*, 263(1), 35-42.
- 425 ³⁴Lau, S., Ng, K. H., & Aziz, Y. F. A. Are volumetric breast density measurements robust enough for routine clinical use?. In *European Congress of Radiology*.
- ³⁵Berg, W. A., Zhang, Z., Lehrer, D., Jong, R. A., Pisano, E. D., Barr, R. G., ... & ACRIN 6666 Investigators. (2012). Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *Jama*, 307(13),
- 430 1394-1404.
- ³⁶Wendie, A., M. D. Berg, and D. Jeffrey. "Combined Screening with Ultrasound and mammography compared to mammography alone in women at elevated risk of breast cancer: results of the first-year screen in ACRIN 6666." *JAMA* 299.18 (2008): 2151-2163.
- ³⁷Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, Morris E, Pisano E, Schnall M, Sener S, et al.: American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007, 57(2):75-89.
- 435 ³⁸Kerlikowske, K., Zhu, W., Hubbard, R. A., Geller, B., Dittus, K., Braithwaite, D., ... & Breast Cancer Surveillance Consortium. (2013). Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. *JAMA internal medicine*, 173(9), 807-816.
- 440 ³⁹Schousboe, J. T., Kerlikowske, K., Loh, A., & Cummings, S. R. (2011). Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Annals of internal medicine*, 155(1), 10-20.
- ⁴⁰Kerlikowske, K., Zhu, W., Tosteson, A. N., Sprague, B. L., Tice, J. A., Lehman, C. D., & Miglioretti, D. L. (2015). Identifying Women With Dense Breasts at High Risk for Interval Cancer: A Cohort Study. *Annals of internal medicine*, 162(10), 673-681.
- 445 ⁴¹Mandelblatt, J. S., Cronin, K. A., Bailey, S., Berry, D. A., de Koning, H. J., Draisma, G., ... & Feuer, E. J. (2009). Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Annals of internal medicine*, 151(10), 738-747.

- 450 ⁴²Braithwaite, D., Zhu, W., Hubbard, R. A., O'Meara, E. S., Miglioretti, D. L., Geller, B., ... & Breast
Cancer Surveillance Consortium. (2013). Screening outcomes in older US women undergoing
multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor
characteristics or false positive rates?. *Journal of the National Cancer Institute*, 105(5), 334-341.
- ⁴³Hartman, K., Highnam, R., Warren, R., & Jackson, V. (2008). Volumetric assessment of breast
tissue composition from FFDM images. *Digital Mammography*, 33-39.
- 455 ⁴⁴Wang, J., Azziz, A., Fan, B., Malkov, S., Klifa, C., Newitt, D., Yitta, S., Hylton, N., Kerlikowske, K.
& Shepherd, J. A. (2013). Agreement of mammographic measures of volumetric breast density to
MRI. *PloS one*, 8(12), e81653.
- ⁴⁵Morrish, O. W., Tucker, L., Black, R., Willsher, P., Duffy, S. W., & Gilbert, F. J. (2015).
Mammographic Breast Density: Comparison of Methods for Quantitative Evaluation. *Radiology*,
460 275(2), 356-365.
- ⁴⁶Schmachtenberg, C., Hammann-Kloss, S., Bick, U., & Engelken, F. (2015). Intraindividual
Comparison of Two Methods of Volumetric Breast Composition Assessment. *Academic radiology*,
22(4), 447-452.
- 465 ⁴⁷van der Waal, D., den Heeten, G. J., Pijnappel, R. M., Schuur, K. H., Timmers, J. M., Verbeek, A.
L., & Broeders, M. J. (2015). Comparing Visually Assessed BI-RADS Breast Density and Automated
Volumetric Breast Density Software: A Cross-Sectional Study in a Breast Cancer Screening Setting.
PloS one, 10(9), e0136667.
- ⁴⁸Broeders, M. J., ten Voorde, M., Veldkamp, W. J., van Engen, R. E., van Landsveld-Verhoeven, C.,
NL't Jong-Gunneman, M., ... & den Heeten, G. J. (2015). Comparison of a flexible versus a rigid
470 breast compression paddle: pain experience, projected breast area, radiation dose and technical image
quality. *European radiology*, 25(3), 821-829.