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


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

<table>
<thead>
<tr>
<th>Title</th>
<th>Impact of errors in recorded compressed breast thickness measurements on volumetric density classification using volpara v1.5.0 software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Article</td>
</tr>
<tr>
<td>URL</td>
<td>This version is available at: <a href="http://usir.salford.ac.uk/38834/">http://usir.salford.ac.uk/38834/</a></td>
</tr>
<tr>
<td>Published Date</td>
<td>2016</td>
</tr>
</tbody>
</table>

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

Ralph Highnam
Matakina Technology Limited, Wellington, New Zealand

Ingrid H. R. Hauge
The Intervention Centre, Rikshospitalet, Oslo University Hospital, Oslo, Norway
Discipline of Medical

Mark F. McEntee
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

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

Judith Kelly
Countess of Chester Hospital NHS Foundation Trust, Chester, UK

Jasmine J. Sarwar
School of Health Sciences, University of Salford, Salford, UK

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

**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.

**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 ±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.

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

**Conclusion:** The current QC standard of ±5mm 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.

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). 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. 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. 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.

There are several methods to measure MD, the most common of which is visual assessment by radiologists using the Breast Imaging Reporting and Data System (BIRADS) scale. This system is prone to inter and intra-reader variability, and thus several computer-assisted methods have been developed including Cumulus and Madena. 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. Automation of this process removes human variability. Several automated systems exist, including AutoDensity, Quantra and Volpara. The latter two use volumetric breast density assessment and has been cleared by the Food and Drug Administrative (FDA) as adjunctive supporting tools. VolparaDensity software uses the raw image and meta data from digital mammograms to calculate average volumetric breast density percentage (VBD%). 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 ±5-10 mm, 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. A typical quality control (QC) tolerance level is up to ±5mm difference between recorded and measured CBT. 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. 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. We noted that techniques exist to detect and correct for compression plate slant. 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). The aim of this study was to investigate the impact of errors
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 ±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. 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. 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. 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.

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 ±15% in 1mm steps. The thickness was adjusted using a software known as DVTk DICOM Editor Tool 3.2.6. 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%, VGD2 = 4.5-7.5% VBD%, VDG3 = 7.5-15.5% VBD%, and VDG4 >15.5% VBD%. 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 (1mm/60mm x100=1.667=2%). The maximum error in recorded CBT was 15% in all cases.
II.D. Statistics
Statistical analyses were conducted using the statistical package R (R for Mac OS, version 3.0.2 GUI 1.62 Snow Leopard build (6558)).\textsuperscript{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 ($V_B$)
3. a function of CBT and fibroglandular volume ($V_f$)
4. a function of CBT, breast volume and fibroglandular volume
III. RESULTS
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).
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
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).

<table>
<thead>
<tr>
<th>Changes in VBD% by error in compressed breast thickness (CBT)</th>
<th>5 mm</th>
<th>15 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean change</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>Mean change when decreasing</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Largest change when decreasing</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Mean change when increasing</td>
<td>-0.4</td>
<td>-0.7</td>
</tr>
<tr>
<td>Largest change when increasing</td>
<td>-2.0</td>
<td>-2.3</td>
</tr>
</tbody>
</table>

III.B. Volpara density grade
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).
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).

<table>
<thead>
<tr>
<th>Changes in VBD% by error in compressed breast thickness (CBT)</th>
<th>VDG1</th>
<th>VDG2</th>
<th>VDG3</th>
<th>VDG4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 mm</td>
<td>15 %</td>
<td>5 mm</td>
<td>15 %</td>
</tr>
<tr>
<td>Mean change</td>
<td>0.2</td>
<td>0.4</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>Largest change when decreasing</td>
<td>0.3</td>
<td>1.0</td>
<td>0.6</td>
<td>1.0</td>
</tr>
<tr>
<td>Largest change when increasing</td>
<td>-0.2</td>
<td>-0.4</td>
<td>-0.5</td>
<td>-0.7</td>
</tr>
</tbody>
</table>

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.
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 deviation for 15% change in recorded CBT.

<table>
<thead>
<tr>
<th>Change in VDG</th>
<th>5mm error in compressed breast thickness (CBT)</th>
<th>15% error in compressed breast thickness (CBT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n=6) Average change in VBD% (SD)</td>
<td>(n=20) Average change in VBD% (SD)</td>
</tr>
<tr>
<td>from 1 to 2</td>
<td>1 0.3 (N/A)</td>
<td>4 0.5 (0.27)</td>
</tr>
<tr>
<td>from 2 to 1</td>
<td>- -</td>
<td>2 0.3 (0.01)</td>
</tr>
<tr>
<td>from 2 to 3</td>
<td>2 0.4 (0.28)</td>
<td>8 0.6 (0.23)</td>
</tr>
<tr>
<td>from 3 to 2</td>
<td>1 0.3 (N/A)</td>
<td>3 0.5 (0.19)</td>
</tr>
<tr>
<td>from 3 to 4</td>
<td>2 0.8 (0.2)</td>
<td>3 0.9 (0.17)</td>
</tr>
<tr>
<td>from 4 to 3</td>
<td>- -</td>
<td>- -</td>
</tr>
</tbody>
</table>
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 ±15% error to the recorded CBT and 11.5 % (n= 6) had a change in VDG within the QC standard prescribed error of ±5mm. 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 (r²=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 ±5mm 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 & Zwiggelaar 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. 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), and this in turn would affect the VBD. Feng et al. 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. 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 al. 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. 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. From this point it is safe to anticipate that in a clinical setting the AEC would affect our results.
The use of MD information has not yet been standardized. The ACRIN 6666 trial confirms the utility of using ultrasound and MRI for women with densest breasts, and additional imaging for women with high MD and higher risk of developing cancer has been suggested. Women with lower familial risk, no genetic markers and lower mammographic density could benefit from less frequent imaging and no additional imaging outside the standard mammographic screening program. 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.

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 UK; 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 manufacturers ranges between ±5-10 mm. 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. If a woman is misplaced from low density 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.

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 ±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 having ±5mm (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.

Our study was solely based on the results of the Volpara software. Both Volpara and Quantra 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 volume. However, the systems have some differences such as internal calibration and correction for compression paddle height and tilt. 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. Quantra has shown higher values of total volume of fibroglandular tissue and VBD, while Volpara has shown higher total breast volume. As the automated 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.
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. 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 12 mm, showing that the current QC standard of 5 mm 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 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 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.

Our study included 52 mammograms. Sample size estimation demonstrated that in order to detect an effect size of 8 mm 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 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
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 ±5mm 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 ±5mm for assessing mammographic density should be reconsidered.
References:


