

Factors Affecting Agreement Between Breast Density Assessment using Volumetric Methods and Visual Analogue Scales

Lucy Beattie¹, Elaine Harkness^{2,3,4*}, Megan Bydder³, Jamie Sergeant^{2,5,6}, Anthony Maxwell^{2,3,4,8}, Nicky Barr³, Ursula Beetles³, Caroline Boggis³, Sara Bundred³, Soujanya Gadde³, Emma Hurley³, Anil Jain³, Elizabeth Lord³, Valerie Reece³, Mary Wilson³, Paula Stavrinou^{3,4}, Gareth Evans^{3,4,7,8}, Tony Howell^{3,4,8}, Susan Astley^{2,3,4,8}

¹ Manchester Medical School, University of Manchester, Stopford Building, Oxford Road, Manchester, M13 9PT

² Centre for Imaging Sciences, Institute of Population Health, University of Manchester, Stopford Building, Oxford Road, Manchester, M13 9PT

³ Genesis Breast Cancer Prevention Centre, University Hospital of South Manchester NHS Trust, Wythenshawe, Manchester M23 9LT, UK

⁴ The University of Manchester, Manchester Academic Health Science Centre, University Hospital of South Manchester NHS Foundation Trust, Wythenshawe, Manchester M23 9LT, UK

⁵ Arthritis Research UK Centre for Epidemiology, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester M13 9PT, UK

⁶ NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre

⁷ Manchester Centre for Genomic Medicine, Central Manchester Foundation Trust, St. Mary's Hospital, Oxford Road, Manchester M13 9WL, UK

⁸ Manchester Breast Centre, Manchester Cancer Research Centre, University of Manchester, Christie Hospital, Withington, Manchester, M20 4BX, UK

Elaine.F.Harkness@manchester.ac.uk

Abstract. Mammographic density in digital mammograms can be assessed visually or using automated volumetric methods; the aim in both cases is to identify women at greater risk of developing breast cancer, and those for whom mammography is less sensitive. Ideally all methods should identify the same women as having high density, but this is not the case in practice. 6422 women were ranked from the highest to lowest density by three methods: QuantraTM, VolparaTM and visual assessment recorded on Visual Analogue Scales. For each pair of methods the 20 cases with the greatest agreement in rank were compared

with the 20 with the least agreement. The presence of microcalcifications, skin folds, suboptimally positioned inframammary folds, and whether or not the nipple was in profile were found to affect agreement between methods ($p < 0.05$). Careful positioning during mammographic imaging should reduce discrepancy, but a greater understanding of the relationship between methods is also required.

Keywords: Digital, mammogram, breast density, agreement, radiographic features.

1 Introduction

Breast density is usually assessed in X-ray mammograms. It varies greatly between women and is influenced by a number of factors including genetics, menopausal status, age and body mass index (BMI) [1]. There are several methods for measuring breast density including subjective area-based methods such as Visual Analogue Scales (VAS), and Breast Imaging and Reporting Data System (BI-RADS) categories for breast density [1,2]. More recently objective volumetric methods have been developed including Quantra™ [3] and Volpara™ [4]. The exact methods for quantifying density used by the manufacturers of Quantra™ and Volpara™ are commercially sensitive, but it is known that different calibration methods are employed, and that the contribution of the skin is disregarded when calculating density measures for Volpara™ therefore we expect the results for Volpara™ to be lower than those for Quantra™. The aim of this study was to identify differences in breast density assessments by Quantra™, Volpara™ and VAS, and to seek to identify patient characteristics, mammographic features and imaging parameters associated with these differences.

2 Methods

Women were recruited from the Predicting Risk of Cancer At Screening (PROCAS) study [3]. PROCAS includes approximately 50,000 women invited for routine breast screening who consented to participate and provided additional information on risk factors for breast cancer by completing a 2-page questionnaire at the time of screening. Women were included if they had density measurements from Quantra™ version 1.3 and Volpara™ version 1.4.0, and from VAS as assessed by two independent readers (VAS1 and VAS2) from a pool of consultant radiologists and advanced radiographic practitioners. VAS forms consisted of four 10cm lines (one for each mammographic view) marked 0% and 100% at each end. Readers place a vertical mark on the line to indicate their assessment of percentage density for each view. Women with missing data for BMI, or with BMI values out with the range 16 to 60 were excluded.

For each measurement method (Quantra™, Volpara™, VAS1, VAS2, VAS average (the average of VAS1 and VAS2)) densities were ranked in ascending order. The range of density measures is different for each method, therefore calculating the dif-

ference in breast density between methods would not necessarily identify the most discrepant cases. On the other hand, discrepancies in ranking enable the identification of subjects where one method assigned a high density score and another method a low density score for the same case. A similar method was used to find the least discrepant cases for each of the pairs of methods. Differences in rank between each pair of methods were sorted in descending order and the cases with the 20 largest and smallest differences for each pair of methods (Quantra™ vs Volpara™; Quantra™ vs VAS average; Volpara™ vs VAS average; and VAS1 vs VAS2) were selected. Where there were ties in the differences in rank between methods these were included, so some groups contain more than 20. Mammograms for each of these women were retrieved and information was extracted by one film reader using a pro-forma.

The pro-forma gathered information on: patient details (age, body mass index); mammographic abnormalities (masses, calcification, distortion, asymmetry, previous breast surgery, previous biopsy (including markers from biopsy)); density results (VAS1, VAS2, Quantra™, Volpara™); positioning (pectoralis muscle – optimal/suboptimal, inframammary folds (IMFs) – optimal/suboptimal, nipple in profile, skin folds, image blurred); radiographic parameters (tube voltage(kV), tube current (mA), breast compression thickness (mm), compression force (N)); and other parameters (visible skin texture, objects in film, non-uniform glandular tissue distribution, prominent vascular markings, prominent lymph nodes). T-tests were carried out for continuous data and Chi-square tests were carried out for categorical data for each of the pairs of methods. Results were considered significant if the p-value was ≤ 0.05 .

3 Results

In total, 6422 women met the entry criteria and had density readings for all three methods. The mean age was 59.3 years (SD 6.7) and mean BMI was 27.9 kg/m² (SD 5.5). Mean densities were 6.8 (SD 3.7), 15.5 (SD 5.7), 29.6 (SD 17.3) and 26.9 (SD 20.0) for Volpara™, Quantra™, VAS1 and VAS2 respectively. Figure 1 shows the level of agreement between the ranks for each pair of methods as well as the least and most discrepant cases selected for the study.

Comparison between Quantra™ and Volpara™.

When comparing Quantra™ and Volpara™, presence of calcification was significantly different between the two measurement methods with those in the most discrepant group having more calcifications than those in the least discrepant group. There were also significantly more suboptimally imaged IMFs in the most discrepant group (Table 1). In addition, there was a significant difference in age between the most and least discrepant groups ($p=0.036$) with those in the most discrepant group being older (mean 61.5) than those in the least discrepant group (mean 56.6).

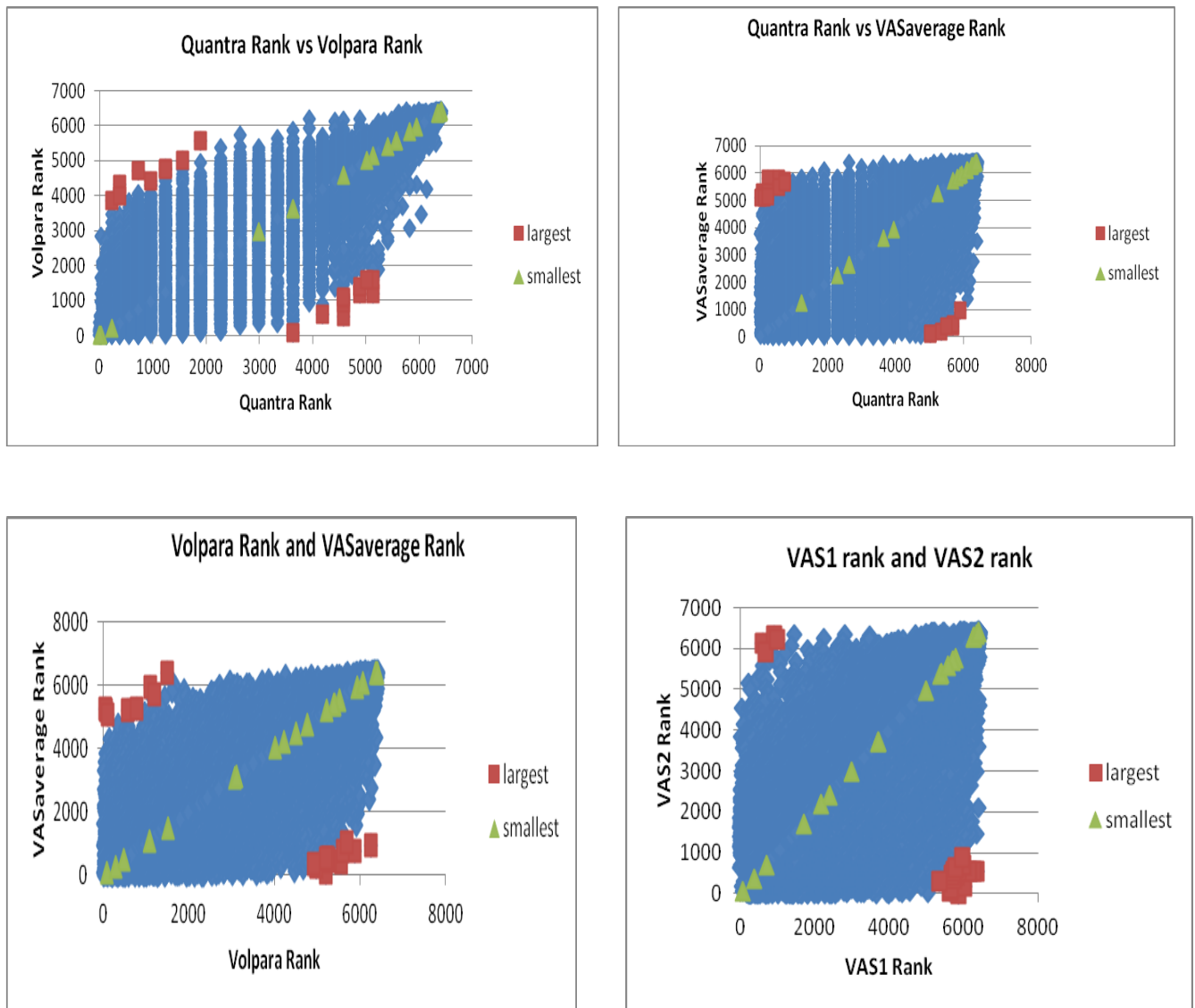


Figure 1: Ranks for each pair of methods (eg. Quantra™ with Volpara™). Those selected for the current study are highlighted- least discrepant cases shown in green and most discrepant cases in red.

Table 1: Comparisons between Quantra™ and Volpara™

Image feature		Most discrepant		Least discrepant		Chi-Square	p-value
		n	%	n	%		
Calcification	No	16	80	21	100	4.65	0.031
	Yes	4	20	0	0		
RMLO Inframammary Fold	Optimal	8	40	15	71	4.11	0.043
	Suboptimal	12	60	6	29		
LMLO Inframammary Fold	Optimal	8	40	18	86	9.23	0.002
	Suboptimal	12	60	3	14		

Comparison between Quantra™ and VAS average

For the comparisons between Quantra™ and average VAS scores, the nipple was not in profile for the LCC view significantly more in the most discrepant group compared to the least discrepant group (Table 2). Similarly, skin folds for the LMLO view were present significantly more in the most discrepant group. Breast thickness (mm) after compression was significantly higher in the most discrepant group for three out of the four mammographic views (RMLO, RCC and LCC) when comparing the difference between Quantra™ and average VAS scores (Table 3). There were no other significant results between Quantra™ and VAS .

Table 2: Comparisons between Quantra™ and VAS average

Image feature		Most discrepant		Least discrepant		Chi-Square	p-value
		n	%	n	%		
LCC nipple in profile	No	6	29	1	4	5.08	0.02
	Yes	15	71	23	96		
LMLO skin folds	Present	6	29	1	4	5.08	0.02
	Absent	15	71	23	96		

Table 3: Comparison of breast thickness (mm) between Quantra™ and average VAS

Projection	Mean Difference (most – least discrepant)	Std. Error Difference	95% Confidence Interval		t	p-value
			Lower	Upper		
RMLO	9.98	4.55	0.81	19.16	2.19	0.03
LMLO	9.52	4.86	-0.29	19.33	1.96	0.06
RCC	9.72	4.07	1.52	17.92	2.39	0.02
LCC	9.05	4.13	0.72	17.38	2.19	0.03

Comparison between Volpara™ and average VAS

The number of women with suboptimal IMFs, for left and right MLO views, were significantly higher in the most discrepant group compared to the least discrepant group. The nipple was also less often in profile, for the RCC view, for the most discrepant group. There were significantly more skin folds in the least discrepant group than the most discrepant group (Table 4). There were also significant differences between compression force for the most and least discrepant groups. Compression force (N) was greater for the most discrepant group for three (RMLO, RCC and LCC) out of the four mammographic views (Table 5).

Table 4: Comparison between Volpara™ and the average VAS scores

Image feature		Most discrepant		Least discrepant		Chi Square	p-value
		n	%	n	%		
RMLO Inframammary Fold	Optimal	8	40	16	73		
	Suboptimal	12	60	6	27	4.58	0.03
LMLO Inframammary Fold	Optimal	8	40	16	73		
	Suboptimal	12	60	6	27	4.58	0.03
RCC nipple in profile	no	5	25	0	0		
	yes	15	75	22	100	6.24	0.01
RMLO skin folds	Present	2	10	10	45		
	Absent	18	90	12	55	6.45	0.01

Table 5: Comparison of compression force (N) between Volpara™ and average VAS

Projection	Mean Difference (most – least discrepant)	Std. Error Difference	95% Confidence Interval		t	p-value
			Lower	Upper		
RMLO	20.77	7.88	4.84	36.71	2.64	0.01
LMLO	12.18	7.98	-3.94	28.30	1.53	0.13
RCC	15.59	7.38	0.67	30.51	2.11	0.04
LCC	26.00	7.09	11.67	40.33	3.67	0.001

Table 6: Difference in tube current (mAs) between VAS1 and VAS2

Projection	Mean Difference (most – least discrepant)	Std. Error Difference	95% Confidence Interval		T	p-value
			Lower	Upper		
RMLO	-14.76	7.25	29.41	-0.11	2.04	0.05
LMLO	-16.29	10.60	37.72	5.14	0.54	0.13
RCC	22.09	26.03	0.51	74.69	0.85	0.40
LCC	-7.72	3.36	14.51	-0.93	2.30	0.03

Comparison between VAS1 and VAS2

There was a statistically significant difference between the tube current recorded for the RMLO and LCC views, with the mean being significantly higher in the least discrepant group. There were no other significant results between VAS1 and VAS2.

4 Discussion

When comparing the different methods for density measurement in the most and least discrepant cases there were a number of interesting findings. Calcification and IMFs were found to be present more often in the most discrepant group when comparing Quantra™ and Volpara™, implying that the two technologies do not agree. How these are dealt with by the computer software designed to measure volumetric density is not publicly known. Those in the most discrepant group also tended to be older than those in the least discrepant group suggesting that volumetric methods may find it more difficult to agree on density of older women's mammograms. Generally, the volume of fibroglandular tissue decreases with age, therefore it is possible that volumetric technologies find smaller volumes of fibroglandular tissue more difficult to interpret, although interestingly, there were no such differences between volumetric methods and VAS density results.

When comparing densities between Quantra™ and VAS average considerably fewer nipples were in profile in the most discrepant group than the least discrepant group for the LCC view. In some images the automated method may have interpreted the nipple as fibroglandular tissue whereas VAS readers would recognise this as a positioning issue. However, this was not found in the three other mammographic views. There was a similar finding for the presence of skin folds. Compression thickness was also significantly different for three mammographic views when Quantra™ and VAS were compared. Compressing the breast minimises movement and spreads the fibroglandular tissue, and dense tissue may be more difficult to assess accurately in a projection image when less compression is used. Interestingly, differences in compression force were not significant in any of the views when comparing these two methods.

Like the comparisons between Quantra™ and VAS, there were a number of significance differences between Volpara™ and VAS for IMFs, nipple not in profile and skin folds. Compression force was found to be higher in the most discrepant group for all four views suggesting less agreement when a larger force is applied.

A limitation of the current study was that density methods were measured across all four views, however the assessment of positioning and information extracted on radiographic parameters from the DICOM headers was for individual mammographic views. It would be interesting to see whether the discrepancies in density are specific to individual mammographic views. Furthermore, assessment of image positioning relied on the interpretation of a single film reader. However images were anonymised and assessed in random order so the reader was blind to which group they came from.

5 Conclusions

This study looked at different methods of measuring breast density: Quantra™, Volpara™ and VAS. There was significantly more calcification and suboptimal IMFs in the most discrepant group when comparing Quantra™ and Volpara™. We also found significant results when comparing the computerised methods with VAS. There were more skin folds in the most discrepant group when comparing Quantra™ and VAS, and the compressed breast thickness was also significantly larger in the most discrepant group. Comparing Volpara™ and VAS found that suboptimal IMFs contributed to disagreement in breast density. Compression force was also significantly higher in the most discrepant group.

At present, there are important differences between different methods of breast density assessment. A greater understanding of the reasons behind such differences coupled with careful radiographic technique should lead to a reduction in these discrepancies.

Acknowledgements

We acknowledge the support of the National Institute for Health Research (NIHR) and the Genesis Prevention Appeal for their funding of the PROCAS study. We would like to thank the women who agreed to take part in the study, the study radiologists and advanced radiographer practitioners for VAS reading, the radiographers involved in the Greater Manchester Breast Screening Programme and the study staff for recruitment and data collection. We also thank Hologic Inc. for providing Quantra™ and Matakina Technology Limited for providing Volpara™. This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research programme (reference number RP-PG-0707-10031: “Improvement in risk prediction, early detection and prevention of breast cancer”). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

6 References

1. Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MR, et al. Heritability of mammographic density, a risk factor for breast cancer. *N Engl J Med* 2002 Sep 19;347(12):886-894.
2. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic Density and the Risk and Detection of Breast Cancer. *N Engl J Med* 2007; 356(3):227-236.
3. Hologic- Quantra Volumetric Assessment. 2012; Available at: <http://www.hologic.com/en/breast-screening/volumetric-assessment/>.
4. Volpara Density. 2013; Available at: <http://www.volparadensity.com/>
5. Evans G. PROCAS study. 2011 [16/06/2013]; Available from: <http://www.uhsm.nhs.uk/research/Pages/PROCASstudy.aspx>.
6. Sergeant, J.C., Warwick, J., Gareth Evans, D., Howell, A., Berks, M., Stavrinou, P., Sahin, S., Wilson, M., Hufton, A., Buchan, I., Astley, S.M.: Volumetric and Area-Based Breast Density Measurement in the Predicting Risk Of Cancer At Screening (PROCAS) Study. In: Maidment, A.D.A., Bakic, P.R., Gavenonis, S. (eds.) *IWDM 2012. LNCS*, vol. 7361, pp. 228–235. Springer, Heidelberg (2012)