Impact of quantitative breast density on experienced radiologists' assessment of mammographic breast density

Poster No.: C-1281

Congress: ECR 2015

Type: Scientific Exhibit

Authors: K. Schilling, J. The, S. Griff, L. Oliver, R. Mahal, M. Saady, M. V. Velasquez; Boca Raton/US

Keywords: Cancer, Observer performance, Mammography, Breast

DOI: 10.1594/ecr2015/C-1281

Aims and objectives

The American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) is the predominant method for assessing mammographic density in the clinic. When this study was conducted, we were using the 4^{th} Edition of BI-RADS. This method involves radiologists visually assigning the breasts into one of four categories [1]:

1. The breast is almost entirely fat (<25% glandular)
2. There are scattered areas of fibroglandular densities (approximately 25-50% glandular)
3. The breast 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)

Breast density is a strong, independent risk factor for developing breast cancer, and increasing breast density also decreases the sensitivity of mammography [2]. Despite the significant implications of breast density on screening mammography and the requirement in many states to inform women there is considerable inter- and intra-observer variability in radiologists' determination of breast density [3, 4]. Even amongst American Board of Radiology examiners, a recent study demonstrated wide variation in the BI-RADS density assessment between them, with kappa scores ranging from 0.347 to 0.665 [5]. This study also demonstrated that technical factors, such as the x-ray system vendor, can influence visual breast density assessment.

To improve the accuracy and reproducibility of breast density assessment, several new methods have been developed, each with their own associated benefits and drawbacks. For example, semi-automated methods such as Cumulus, can predict the risk of developing breast cancer, but are too labour-intensive for widespread clinical use [6]. Fully-automated area-based methods, such as ImageJ and AutoDensity, are still being validated clinically, but are limited by the fact that they are measuring a 3-dimensional phenomenon, from a 2-dimensional projection [6, 7]. Fully-automated volumetric-based methods that measure the physical amount of fibroglandular tissue in the breast are reproducible, proven to be associated with breast cancer risk, FDA cleared and slowly coming into clinical use [8]. We sought to study whether the use of fully-automated volumetric breast density (VBD) software improves the visual BI-RADS inter-reader agreement between radiologists at our clinic.

Methods and materials

The methods used for this study are outlined in **Figure 1**. Eight experienced breast imaging radiologists, from a single facility, visually assigned 100 digital screening mammographic studies into one of four BI-RADS breast composition categories. 12 studies comprised women with breast implants and were excluded from the study, leaving 88 studies in the final analyses. For each pair of readers (i.e. 28 pairs), the inter-reader agreement was assessed using Cohen's kappa coefficient (k).

The corresponding raw (for processing) images were processed using fully-automated quantitative software (VolparaDensity™ v1.4, Volpara Solutions, Wellington, New Zealand) to generate volumetric breast density and an associated Volpara Density Grade (VDG), equivalent to BI-RADS density category. After a 2-week washout period, all radiologists re-assessed breast density from the same 88 digital mammograms, with the VDG scores from VolparaDensity available during the reading session. As described above, Cohen's kappa coefficient was again used to assess inter-reader agreement for each pair of readers.

To assess whether the inter-reader agreement improved with the use of automated density software as an aid, the Wilcoxon rank-sum test was used to compare distributions of kappa scores without and with VolparaDensity..

Images for this section:

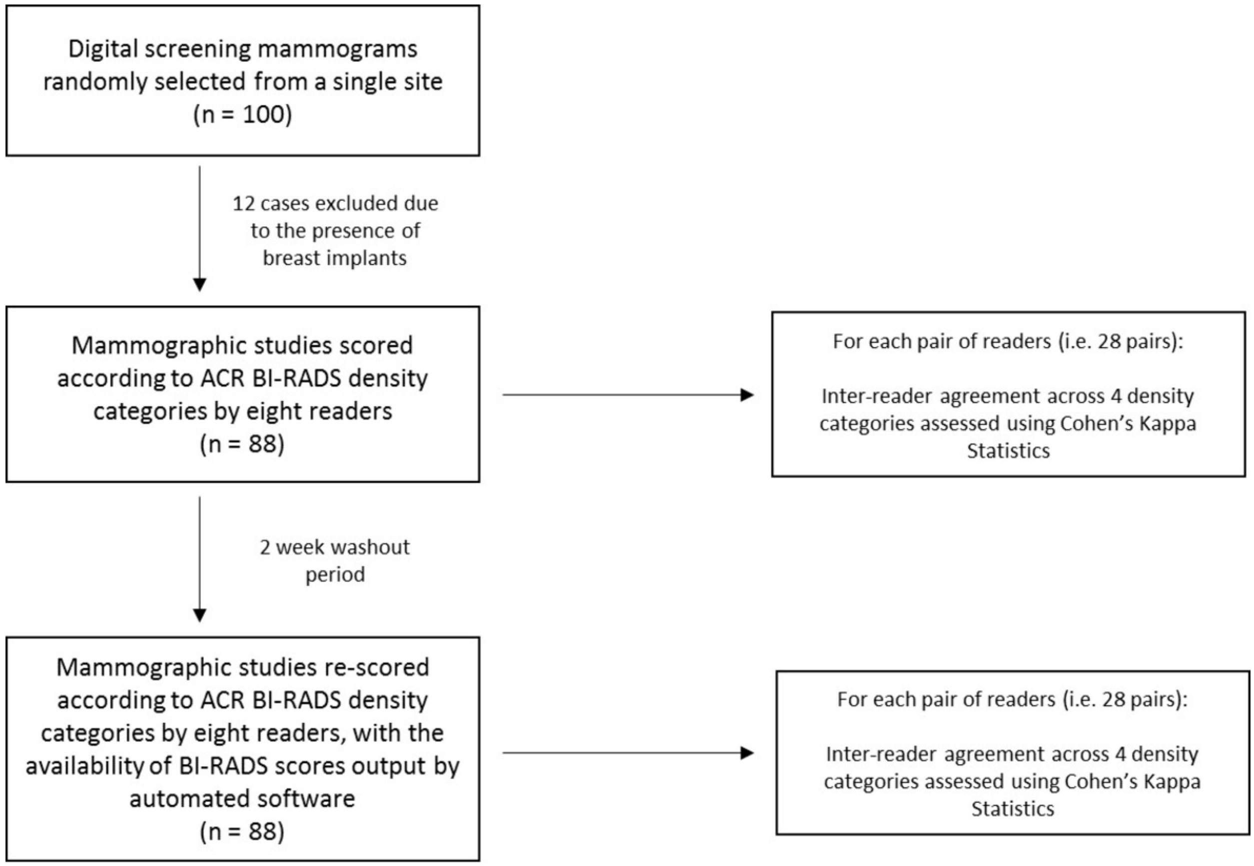

Fig. 1: Flow diagram outlining the methodology and analyses used in the present study.

Results

The overall visual BI-RADS distributions for each reader and the VDG distributions are shown in **Table 1**, without (top panel) and with (bottom panel) the aid of Volpara. Without the aid of Volpara, there was considerable variability in the proportion of women assigned into each BI-RADS category. For example, without the software aid, the percentage of women assigned as BI-RADS 1, BI-RADS 2, BI-RADS 3 and BI-RADS 4 ranged from 3.4 - 27.3%, 47.7 - 71.6%, 17.0 - 35.2%, and 0 -6.8%, respectively. When the VDG scores from Volpara were available, the range in the proportion of women allocated into each category by the eight readers appeared to be significantly reduced i.e. 8.0 - 18.2%, 46.6 - 62.5%, 22.7 - 34.1%, and 3.4 - 10.2%, respectively. In comparison, Volpara assigned 18.2, 40.9, 28.4 and 12.5% of women into each BI-RADS category, respectively.

The percentage agreement between each pair of readers, using a four-category density scale is shown in **Table 2**, without (top panel) and with (bottom panel) the aid of Volpara. On average, the agreement between pairs of readers improved by approximately two percentage points when the software aid was used.

Table 3 demonstrates the level of agreement without (top panels) or with (bottom panels) using kappa statistics. The use of quantitative VBD significantly improved inter-reader agreement in radiologists' assessment of breast density (p=0.0374) with a mean kappa without VDG of 0.5664 and with VDG of 0.6266.

Although the above results demonstrated tha t use of a density software aid can improve the agreement between readers, the effectiveness of the aid depends largely on the willingness of the reader to amend their scores based on an objective measure of density. **Table 4** outlines the intra-reader agreement for each individual reader without and with the software aid. Although there is always some inherent intra-reader variability, these results suggest that some readers are fairly confident in their own readings , while some are more amenable to changing their density readings. For example, having access to the objective scores resulted in reader 4 only changing their final density assessment in less than 10% of cases, whereas reader 3 changed their density assessment in over 30% of cases.

Images for this section:

BI-RADS distributions (%) without Volpara									
	R1	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**	**Volpara**
BI-RADS 1	9.1	27.3	3.4	8.0	15.9	9.1	10.2	19.3	18.2
BI-RADS 2	71.6	47.7	76.1	50.0	56.8	56.8	65.9	58.0	40.9
BI-RADS 3	17.0	20.5	18.2	35.2	22.7	31.8	19.3	22.7	28.4
BI-RADS 4	2.3	4.5	2.3	6.8	4.5	2.3	4.5	0	12.5

BI-RADS distributions (%) with Volpara									
	R1	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**	**Volpara**
BI-RADS 1	10.2	18.2	11.4	11.4	9.1	10.2	12.5	8.0	18.2
BI-RADS 2	61.4	47.7	52.3	46.6	53.4	58.0	59.1	62.5	40.9
BI-RADS 3	22.7	23.9	28.4	34.1	28.4	27.3	22.7	26.1	28.4
BI-RADS 4	5.7	10.2	8.0	8.0	8.0	4.5	5.7	3.4	12.5

Table 1: Showing the BI-RADS distributions assessed visually by each of the eight readers, or using the Volpara software. Top and bottom panels show the distributions without and with the aid of Volpara software density scores, respectively.

Percentage agreement (%) without Volpara - 4 density categories								
	R1	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**
R1		72.73	81.82	63.64	81.82	77.27	84.09	79.55
R2			61.36	59.09	72.73	70.45	79.55	82.95
R3				67.05	73.86	78.41	80.68	71.59
R4					72.73	81.82	75	62.5
R5						81.82	81.82	78.41
R6							84.09	73.86
R7								80.68
R8								
Mean	**75.41**							

Percentage agreement (%) with Volpara - 4 density categories								
	R1	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**
R1		65.91	68.18	68.18	68.97	73.86	76.14	77.27
R2			79.55	76.14	85.06	70.45	81.82	73.86
R3				73.86	83.91	75	84.09	78.41
R4					80.46	78.41	76.14	75
R5						79.31	87.36	78.16
R6							81.82	82.95
R7								85.23
R8								
Mean	**77.34**							

Table 2: Showing the percentage agreement between each pair of readers using the four BI-RADS categories, without (top panel) and with (bottom panel) the aid of Volpara software.

Cohen's Kappa Statistic – Without Volpara (4 density categories)								
	R1	R2	R3	R4	R5	R6	R7	R8
R1		0.5436	0.5674	0.3656	0.6627	0.5714	0.6719	0.6132
R2			0.3441	0.3844	0.5718	0.5368	0.6679	0.7269
R3				0.4019	0.497	0.5733	0.5787	0.4439
R4					0.5602	0.6944	0.5771	0.39
R5						0.6915	0.6772	0.633
R6							0.7127	0.55
R7								0.6515
R8								
Mean	0.5664							

Cohen's Kappa Statistic – With Volpara (4 density categories)								
	R1	R2	R3	R4	R5	R6	R7	R8
R1		0.4576	0.4684	0.4872	0.4746	0.5409	0.5811	0.5845
R2			0.6872	0.6425	0.7686	0.5348	0.7138	0.5794
R3				0.5919	0.7409	0.5863	0.7383	0.6324
R4					0.6925	0.6528	0.6203	0.5888
R5						0.6535	0.7878	0.6225
R6							0.6866	0.6939
R7								0.7359
R8								
Mean	0.6266							

Table 3: Showing the individual and mean kappa values across all 28 pairs of readers, without (top panel) and with (bottom panel) the aid of Volpara software, using the four-category BI-RADS classification system.

	Intra-reader Analysis								
		R1	R2	R3	R4	R5	R6	R7	R8
R1	Kappa (without Volpara)	0.4889							
	Kappa (with Volpara)	0.5227							
	Percentage agreement (%)	73.86							
R2	Kappa (without Volpara)		0.5755						
	Kappa (with Volpara)		0.6893						
	Percentage agreement (%)		71.59						
R3	Kappa (without Volpara)			0.4158					
	Kappa (with Volpara)			0.5165					
	Percentage agreement (%)			68.18					
R4	Kappa (without Volpara)				0.8563				
	Kappa (with Volpara)				0.8863				
	Percentage agreement (%)				90.91				
R5	Kappa (without Volpara)					0.6427			
	Kappa (with Volpara)					0.7211			
	Percentage agreement (%)					78.16			
R6	Kappa (without Volpara)						0.6632		
	Kappa (with Volpara)						0.7171		
	Percentage agreement (%)						80.68		
R7	Kappa (without Volpara)							0.8145	
	Kappa (with Volpara)							0.8509	
	Percentage agreement (%)							89.77	
R8	Kappa (without Volpara)								0.6367
	Kappa (with Volpara)								0.6751
	Percentage agreement (%)								79.55

Table 4: Showing the intra-reader agreement for each individual reader. Kappa values and percentage agreements are indicative of the intra-reader agreement without or with the Volpara software aid.

Conclusion

Consistency in the determination of breast density is important for clinical decision-making regarding breast cancer risk assessment and adjunctive imaging in women with dense breasts.

Automated breast density software significantly improved inter-reader agreement of experienced radiologists' assessment of mammographic breast density, when using the ACR BI-RADS 4-category classification system. The effectiveness of any particular software aid in improving the standardization of breast density assessment in the clinic, will depend, in part, on clinicians accepting and familiarizing themselves with objective measurements of breast density. Reassuringly, it appears most readers in this study accepted and used the automated scores to improve their readings.

Personal information

KATHY SCHILLING, M.D., JULIETTE THE, M.D., SIMONE KAHN GRIFF, M.D., LAWRENCE H. OLIVER, M.D., RAVINDER SINGH MAHAL, M.D., MATTHEW SAADY M.D., MARIA VICTORIA VELASQUEZ, M.D.

Boca Raton Regional Hospital, Boca Raton, Florida 33486

References

[1] American College of Radiology. Breast Imaging Reporting and Data System® (BI-RADS®) 4. Reston, Va: American College of Radiology; 2003

[2] Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, Jong RA, Hislop G, Chiarelli A, Minkin S, Yaffe M. Mammographic density and the risk and detection of breast cancer. NEJM. 2007; 356:227-36

[3] Fujita H, Hara T, Muramatsu C (Eds): IWDM 2014, LNCS 8539, pp. 239-244, 2014

[4] Ciatto S, Houssami N, Apruzzese A, Bassetti E, Brancato B, Carozzi F, Catarzi S, Lamberini MP, Marcelli G, Pellizzoni R, Pesce B, Risso G, Russo F, Scorsolini A.

Categorizing breast mammographic density: intra- and inter-reader reproducibility of BI-RADS density categories. Breast. 2005; 14:269-275

[5] McEntee M and Damases C. Mammographic density measurement: a comparison of automated volumetric density measurement to BIRADS, Proc SPIE 9037, Medical Imaging 2014 Image Perception, Observer Performance and Technology Assessement, 90370T (March 11, 2014); doi: 10.1117/12.2042966

[6] Sovio U, Li J, Aitken Z, Humphreys K, Czene K, Moss S, Hall P, McCormack V, dos-Santos-Silva I. Comparison of fully and semi-automated area-based methods for measuring mammographic density and predicting breast cancer risk. Br J Cancer. 2014; 110:1908-16

[7] Nickson C, Arzhaeva Y, Aitken Z, Elgindy T, Buckley M, Li M, English DR, Kavanagh AM. AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. Breast Cancer Res. 2013; 15:R80

[8] Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, Vinnicombe S, Allen S, dos-Santos-Silva I. Digital mammgographic density and breast cancer risk: a case-control study of six alternative density assessment methods.