Received:
16 December 2015

Revised:
6 February 2016

Accepted:
15 February 2016

<http://dx.doi.org/10.1259/bjr.20151057>

Cite this article as:

Ekpo EU, McEntee MF, Rickard M, Brennan PC, Kunduri J, Demchig D, et al. Quantra™ should be considered a tool for two-grade scale mammographic breast density classification. *Br J Radiol* 2016; **89**: 20151057.

FULL PAPER

Quantra™ should be considered a tool for two-grade scale mammographic breast density classification

^{1,2}ERNEST U EKPO, BSc (Hons), ¹MARK F MCENTEE, PhD, ^{1,3}MARY RICKARD, MD, ¹PATRICK C BRENNAN, PhD, ³JYOTSNA KUNDURI, MD, ¹DELGERMAA DEMCHIG, MD and ¹CLAUDIA MELLO-THOMS, PhD

¹Discipline of Medical Radiation Sciences in the Faculty of Health Sciences and the Brain and Mind Research Institute, The University of Sydney, Australia

²Department of Radiography and Radiology, University of Calabar, Calabar, Nigeria

³Sydney Breast Clinic, Sydney, NSW, Australia

Address correspondence to: Mr Ernest U Ekpo

E-mail: eeep9437@uni.sydney.edu.au

Objective: To assess the agreement between Quantra™ (Hologic Inc., Bedford, MA) and Breast Imaging Reporting and Data Systems (BI-RADS®) and the performance of Quantra at reproducing BI-RADS mammographic breast density (MBD) assessment.

Methods: MBD assessment was performed using Quantra and BI-RADS. BI-RADS assessment was performed in two phases (1314 and 292 cases, respectively). Kappa was used to assess the interreader agreement and the agreement between Quantra and BI-RADS, and receiver-operating characteristics analysis was used to assess the performance of Quantra at reproducing BI-RADS rating.

Results: Agreement (weighted kappa) between BI-RADS and Quantra in Phase 1 was 0.75 [95% confidence interval (CI): 0.73–0.78] and 0.85 (95% CI: 0.80–0.90) on four- and

two-grade scales, respectively. The corresponding agreement in Phase 2 was 0.79 (95% CI: 0.75–0.84) and 0.84 (95% CI: 0.79–0.87) using the majority report. In Phase 1, Quantra demonstrated 93.2% sensitivity and 86.1% specificity for BI-RADS on a two-grade scale (1–2 vs 3–4). In Phase 2, it demonstrated 91.3% sensitivity and 83.6% specificity on a two-grade scale.

Conclusion: Quantra is limited in reproducing BI-RADS rating on a four-grade scale; however, it highly reproduces BI-RADS assessment on a two-grade scale.

Advances in knowledge: Quantra (v. 2.0) is a poor predictor of BI-RADS assessment on a four-grade scale, but well reproduces BI-RADS rating on a two-grade scale. Therefore, it should be considered a tool for two-grade scale MBD classification.

INTRODUCTION

Breast density is associated with breast cancer risk, interval breast cancer and mammographic sensitivity.^{1,2} It has been shown that 30% reduction in breast cancer mortality is partly attributable to improved knowledge of its risk factors and advances in early detection and treatment strategies.^{3,4} Breast density refers to the proportion of the breast that is composed of fibroglandular tissue⁵ and is related to established breast cancer risk factors such as genetics, hormonal agents and lifestyle characteristics.^{6,7} Most importantly, breast density reduction has been shown to be a prognostic marker of response to adjuvant tamoxifen therapy in post-menopausal females.^{8,9} Therefore, adding breast density information to existing breast cancer risk-prediction models may improve the stratification of breast cancer risk,¹⁰ which may lead to earlier adoption of prevention and control measures. Breast density information can also be used for tailoring clinical decisions such as screening intervals for females who are asymptomatic and the selection of more appropriate imaging pathways such

as ultrasound,¹¹ MRI¹² or digital breast tomosynthesis (DBT)¹³ to enhance the visualization and evaluation of the dense tissue for features of cancer.¹⁴

Currently, breast density notification laws are effective in 22 US states, and the most economically viable and common method of obtaining such information is through mammographic breast density (MBD) assessment.⁵ Qualitative, semi-automated and fully automated approaches are available for acquiring MBD information from mammograms.¹⁵ Qualitative approaches such as Breast Imaging Reporting and Data Systems (BI-RADS®) subjectively assign MBD grades based on the perceived relative proportions of fatty and dense tissue and ductal distribution. Semi-automated techniques involve the user outlining the dense area using computer assistance to calculate percentage MBD. Fully automated area-based methods such as MedDensity and Autodensity measure percentage MBD, whilst others such as Quantra™ (Hologic Inc., Bedford, MA) and Volpara™ (Volpara Solutions, Mākatina Company,

Wellington, New Zealand) perform volumetric breast density measurement.¹⁵ All approaches have shown that breast density is a significant risk factor for breast cancer.^{16,17} However, for breast density information to be used consistently for assessing risk and tailoring screening pathways, it is essential that methods of assessment are reliable and reproducible, as inaccurate calculation could adversely affect the patient outcome and clinical management. The BI-RADS approach commonly used in the clinical practice is subjective and inconsistent.^{18,19} Semi-automated methods require user input and also suffer from subjective variability and add labour-intensive software into the already complex radiological reporting ecosystem.²⁰ There are also concerns that measurement of MBD as percentage area breast density by area-based semi-automated and automated methods is affected by breast size and treatment of the breast as a two-dimensional structure and therefore may not be representative of the quantity of the dense tissue in the breast.²¹ To avoid variability and provide a more consistent measure of the quantity of the dense tissue in the breast, automated volumetric tools such as Quantra have been developed.

Quantra is designed to estimate volumetric breast density and generate pseudo-BI-RADS grades. For this reason, it is important when using Quantra for MBD assessment in clinical practice to measure to what degree its MBD measures compare with radiologist BI-RADS grading. Only two peer-reviewed articles have reported the performance of the Quantra in reproducing radiologists' fourth edition BI-RADS assessment and one on the fifth edition of BI-RADS rating scale.²² Each of these studies are laboratory based and employed a relatively small sample size and have reported different sensitivity, specificity and cut-off values for Quantra in reproducing BI-RADS categories on a two-grade scale (BI-RADS 1–2 vs 3–4), but none have reported Quantra performance on a four-grade scale (BI-RADS 1–4).^{23–25} These studies have recommended further works to assess the performance of Quantra in order to ensure that it is fit for the purpose of clinical MBD assessment. Furthermore, these studies employed older versions of Quantra (v. 1). A recalibrated and upgraded newer version (v. 2.0) has been developed; however, it is unknown how well this new version will perform in MBD assessment. The recommendations for further studies and the deficiencies highlighted above underscore the need for further evaluation of Quantra. Also, BI-RADS is observer dependant,^{18,19} and extraneous variables such as artificiality of the study environment and extent of scrutiny of one's performance influence reader behaviour in a laboratory-type experimental setting (laboratory effect).^{26,27} Since these variables cause discrepancy between results obtained in a laboratory-type environment and real-life clinical situations,^{26,27} it is imperative to assess how Quantra assessment compares with radiologist-assigned BI-RADS categories in a real-life clinical environment. In addition, the association of high-grade breast densities (BI-RADS 3–4) with cancer risk and mammographic masking effect emphasizes the need for further assessment of the cut-off value for the correct differentiation between high- and low-grade BI-RADS categories using Quantra. Therefore, this study aimed to assess the agreement between Quantra and BI-RADS. It also aimed to assess the performance of Quantra in reproducing BI-RADS assessment and the cut-off value for Quantra that best differentiates high- and low-grade BI-RADS categories.

METHODS AND MATERIALS

This study involved 1314 females who underwent screening mammography between March and July 2014. All females gave both oral and written consent for use of their mammograms for research purposes. Mammograms were acquired using the Hologic® Selenia® Dimensions® equipment (Hologic Inc.). The equipment comprised of a tungsten (W) anode and rhodium (Rh) and silver (Ag) filters. All mammograms were acquired in standard craniocaudal and mediolateral oblique projections with automatic image-acquisition parameters [tube potential (kVp), tube current (mAs) and filter type and thickness].

Mammographic breast density assessment

Two methods of MBD measurement were used, automated breast density assessment and qualitative assessment, by radiologists. Automated breast density assessment from the raw projection images was performed using Quantra v. 2.0 (Hologic Inc.). Quantra employs a physical modelling of mammographic systems to calculate volumetric breast density. Volumetric breast density measurement is based on the physical composition of the breast and compressed breast thickness, as well as the information related to the X-ray spectra such as tube potential (kVp), tube current (mAs) and filter type and thickness.^{28,29} Quantra estimates the thickness of the dense tissue above each pixel in each mammographic image and combines these pixel values to compute the total volume of the dense tissue in the breast [V_{fg} (cm³)]. It examines the entire imaged breast silhouette to estimate the entire volume of the breast [V_{bd} (cm³)] and calculates percentage volumetric breast density (V_{bd} %) as a percentage of the estimated dense tissue volume and the total breast volume. The software also calculates the area percentage breast density (A_{bd} %) as the area of the fibroglandular tissue and the total area of the breast. Quantra segments the estimated volumetric breast density to generate fractional quantized breast density (q_{abd}) values for each mammographic view. These q_{abd} estimates from all views are averaged to produce a total q_{abd} value for each patient, which is then used to generate a total quantized breast density (Q_{abd}) value for that patient. Quantra estimates the gross breast density make-up of a patient (total Q_{abd}) by rounding off the total q_{abd} value for that patient to the nearest integer. Briefly, total q_{abd} values are rounded off as follows: ≤1.44 represents a total Q_{abd} value of 1; 1.45–2.44 represents a total Q_{abd} value of 2; 2.45–3.44 represents a total Q_{abd} value of 3; and ≥3.45 represents total a Q_{abd} value of 4. Quantra's Q_{abd} values range from 1 to 4 and are used to map BI-RADS 1–4 categories, respectively. For this study, the total Q_{abd} value estimated by Quantra for each patient was matched to radiologists' BI-RADS assessment for that patient and used to assess the ability of Quantra to reproduce BI-RADS assessment.

Qualitative MBD assessment was performed by radiologists using the fourth edition BI-RADS scheme on a discrete scale (1–4). For this study, BI-RADS breast density assessment was performed in two phases. In the first phase, 1314 cases had their MBD assigned in a single reading during clinical mammography interpretation and were stored in a database; each case was read by 1 radiologist, and in total, 6 breast radiologists were involved across all these cases. All readers were Royal Australian and New

Zealand College of Radiology-certified breast radiologists. Readers were working in the same practice and in the New South Wales breast-screening programme (BreastScreen, NSW, Australia). The characteristics of the readers and the number of cases each reader assessed in the present study are shown in Table 1. These BI-RADS MBD assessment reports were retrieved from the database and matched with the breast density grades assigned by Quantra.

In Phase 2, a subsample of 300 mammograms was selected from the 1314 cases and the MBD grades assigned to these cases in Phase 1 recorded. This subsample consisted of the first consecutive 60, 80, 90 and 70 cases rated BI-RADS 1, 2, 3 and 4, respectively, in Phase 1. The selection was to ensure that there were a substantial number of cases in each category and to fairly represent their prevalence in the original data set in Phase 1. This subsample was assessed by three radiologists (MR, DD and JK) blinded to the results of Quantra assessment and a majority report was generated (consensus of at least two radiologists); two of the radiologists (MR and JK) participated in Phase 1. A repeat assessment of the first consecutive 100 cases in the data set was performed by each radiologist 5 months after the initial assessment.

The agreement between each radiologist and the majority report was assessed by comparing their first readings with the majority report, and interreader agreement was assessed by comparing the first assessments of the radiologists in pairs. Intrareader agreement was assessed by comparing the first and second readings of each radiologist. Data of each reader were grouped into two categories (1–2 and 3–4) to enable the assessment of reader agreement on a two-category scale (1–2 vs 3–4). Also, the BI-RADS assessment in Phase 1 and the majority report in Phase 2 were grouped into a two-grade scale (1–2 and 3–4) to assess the performance of Quantra in reproducing BI-RADS categories on a two-grade scale.

Statistical analysis

The Statistical Package for Social Sciences v. 21 (IBM Corp., New York, NY; formerly SPSS Inc., Chicago, IL) was used for data

analysis. Weighted kappa (κ_w) was also used to calculate the agreement between each radiologist and the majority report in Phase 2 as well as intrareader and interreader agreement. The agreement between Quantra and each reader as well as their majority report was calculated using both simple and weighted kappa. Simple kappa was calculated to show the percentage agreement between Quantra and BI-RADS, whilst weighted kappa was calculated to account for agreement beyond chance expectation. The agreement (κ) between the majority report generated in Phase 2 and the initial BI-RADS reports provided for these cases in Phase 1 was also calculated. Agreement was assessed on a four-grade scale (BI-RADS 1–4) and on a two-grade scale (BI-RADS 1–2 vs 3–4). Kappa was interpreted as designated by Landis and Koch:³⁰ 0.00–0.20 (slight), 0.21–0.40 (fair), 0.41–0.60 (moderate), 0.61–0.80 (substantial) and 0.81–1.00 (almost perfect) agreements. Receiver-operating characteristics (ROC) analysis was also used to assess the performance of Quantra in reproducing BI-RADS assessments in both phases. To assess the performance of Quantra in reproducing BI-RADS on a two-grade scale, BI-RADS 1–2 were grouped together, as were 3–4. This enabled assessment of the sensitivity, specificity and cut-off value for the correct differentiation of high- and low-grade BI-RADS categories using Quantra. The area under the curve (AUC) was used as a figure of merit to assess the overall performance of Quantra in reproducing BI-RADS. AUC was interpreted as described by Fan et al:³¹ 1 (perfect); 0.9–0.99 (excellent); 0.8–0.89 (good); 0.7–0.79 (fair); 0.51–0.69 (poor); and 0.5 (worthless). The University of Sydney institutional review board approved the study (IRB: 2014/481).

RESULTS

The 1314 cases used in Phase 1 comprised of 13.8% ($n = 181$), 30.1% ($n = 396$), 37.8% ($n = 497$) and 18.3% ($n = 240$) for BI-RADS 1, 2, 3 and 4, respectively. MBD assessment readings were not provided for 8 cases in Phase 2, leaving a total of 292 cases that were assessed by all three radiologists. The majority report of the three radiologists in this phase consisted of 9.6% ($n = 28$), 35.3% ($n = 103$), 27.1% ($n = 79$) and 28.1% ($n = 82$) for BI-RADS 1, 2, 3 and 4, respectively. The number of cases classified as BI-RADS 1, 2, 3 and 4 ranged from 24 to 60 (mean = 39.7),

Table 1. Characteristics of readers and number of cases assessed

Readers	Number of years certified	Number of years screen reading mammograms	Number of cases read per week	Number of hours reading per week	Number of cases assessed in the present study
R1	13	13	101–150	21–30	326
R2	20	20	Over 200	16–20	219
R3	3	3	Over 200	5–10	189
R4	20	20	Over 200	16–20	272
R5	19	19	101–150	Over 30	201
R6	35	25	100	2–10	107
	Mean = 18.3	Mean = 16.7	Range from ≥ 100 to >200	Range from ≥ 5 to >30	Total = 1314

R, reader.

80–105 (mean = 90.7), 81–85 (mean = 82.3) and 60–82 (mean = 79.3), respectively. Weighted kappa analysis in Phase 2 demonstrated an almost perfect intrareader agreement (κ), as shown in Table 2.

A substantial to almost perfect agreement was observed between the individual radiologist and the majority report on a four-grade scale from 0.80 (95% CI: 0.76–0.83) to 0.89 (95% CI: 0.84–0.93) and almost perfect agreement on a two-grade scale from 0.82 (95% CI: 0.77–0.87) to 0.90 (95% CI: 0.85–0.94). There was substantial interreader agreement on a four-grade scale from 0.66 (95% CI: 0.62–0.71) to 0.75 (95% CI: 0.70–0.81) and substantial 0.77 (95% CI: 0.73–0.82) to almost perfect 0.89 (95% CI: 0.84–0.93) interreader agreement on a two-grade scale (Table 2).

The agreement (κ_w) between the Phase 1 BI-RADS assessment and the majority report in Phase 2 was 0.78 (95% CI: 0.73–0.85) on a four-grade scale and 0.85 (95% CI: 0.79–0.88) on a two-grade scale.

Quantra percentage volumetric density estimates are converted to pseudo-BI-RADS four-point scale based on preset percentages; for this study, these Quantra percentage volumetric density were 1–7%, 6–22%, 13–29% and 25–44% for BI-RADS 1, 2, 3 and 4, respectively. The distribution of Quantra classification of the Phase 1 and Phase 2 BI-RADS assessment is shown in Figure 1a,b, respectively.

Agreement (κ_w) between BI-RADS and Quantra in Phase 1 was 0.75 (95% CI: 0.73–0.78) and 0.85 (95% CI: 0.80–0.90) on four- and two-grade scales, respectively. The corresponding simple kappa was 0.53 (95% CI: 0.50–0.57) and 0.59 (95% CI: 0.55–0.68). The agreement between BI-RADS majority report and Quantra in Phase 2 is shown in Table 3. The agreement (κ_w) between readers and Quantra ranged from 0.77 (95% CI: 0.72–0.81) to 0.81 (95% CI: 0.76–0.84) on a four-grade scale and from 0.81 (95% CI: 0.74–0.86) to 0.83 (95% CI: 0.79–0.87) on a two-grade scale (Table 3).

Considering the BI-RADS assessment in Phase 1 as the assumed truth, ROC analyses show that on a four-grade scale (BI-RADS 1–4), Quantra demonstrated a sensitivity for 25.4%, 81.1%, 90.7% and 56.7% of BI-RADS 1, 2, 3 and 4, respectively. On a two-grade scale (1–2 vs 3–4), Quantra demonstrated 93.2% sensitivity and 86.1% specificity (Table 4), and a cut-off value of 20% allowed differentiation between high- and low-grade BI-RADS categories.

Restricting the analyses to the majority report in Phase 2, Quantra demonstrated a sensitivity of 35.7%, 91.2%, 88.6% and 50.3% for BI-RADS 1, 2, 3 and 4, respectively. On a two-grade scale (1–2 vs 3–4), it demonstrated 91.3% sensitivity and 83.6% specificity (Table 5), and a cut-off value of 20.5% allowed differentiation between high- and low-grade BI-RADS categories.

DISCUSSION

Since BI-RADS depends on individual radiologist perception of MBD,^{18,19,23} it is intuitive that an automated method such as Quantra, which is modelled with input from radiologists and selectively measures the volume of the dense tissue, may demonstrate greater reproducibility. However, this alone is not sufficient to warrant clinical implementation; automated measures must reflect the clinical opinion of the experienced breast radiologist consistently. In consideration of this, the present work examines how well the automated measures reflect BI-RADS assessment and treats BI-RADS as the “truth”. Simple kappa (a measure of percentage agreement) demonstrated moderate agreement between Quantra and BI-RADS. Weighted kappa, which provides a measure of agreement beyond chance expectation, demonstrated a good level of agreement between these two methodologies on a two-grade scale (1–2 vs 3–4). Similarly, ROC analysis showed that Quantra is limited in reproducing BI-RADS rating on a four-grade scale, but very well reproduces BI-RADS on a two-grade scale (Table 4). The area under the curve (AUC = 0.917; 95% CI: 0.901–0.933) indicates that Quantra is a promising tool for MBD assessment on a two-grade scale in the clinical setting.

Table 2. Weighted kappa (κ_w) analyses [95% confidence interval (CI)] showing intrareader agreement, each reader's agreement with the majority report and their interreader agreement in Phase 2

Readers	Kappa values (95% CI) on a four-grade scale	Kappa values (95% CI) on a two-grade scale
Intrareader (R1)	0.87 (0.83–0.92)	0.91 (0.88–0.95)
Intrareader (R2)	0.86 (0.83–0.91)	0.88 (0.83–0.92)
Intrareader (R3)	0.88 (0.85–0.93)	0.90 (0.87–0.94)
R1 vs majority	0.80 (0.76–0.83)	0.82 (0.77–0.87)
R2 vs majority	0.89 (0.84–0.93)	0.90 (0.85–0.94)
R3 vs majority	0.83 (0.78–0.86)	0.84 (0.79–0.88)
R1 vs R2	0.73 (0.68–0.77)	0.89 (0.84–0.93)
R1 vs R3	0.66 (0.62–0.71)	0.83 (0.79–0.85)
R2 vs R3	0.75 (0.70–0.81)	0.77 (0.73–0.82)
Overall (R1, 2, 3)	0.72 (0.67–0.83)	0.84 (0.78–0.89)

R, reader.

Figure 1. Distribution of Quantra™ (Hologic Inc., Bedford, MA) mammographic breast density (MBD) assessment relative to BI-RADS® in Phase 1 single report (a) and Phase 2 majority report (b).

On a four-grade scale, Quantra demonstrated low sensitivity but high specificity for BI-RADS 1 and 4; it showed considerably high sensitivity for BI-RADS 2 and 3 in both phases of the assessment (Tables 4 and 5). Clinically, the differentiation between BI-RADS 2 and 3 is the most important, as it determines clinical decisions made from MBD assessment such as identification of females at elevated risk of breast cancer.¹⁰ It also guides clinicians in selecting females who may require additional imaging with ultrasound or MRI, as well as the next screening round for females who are asymptomatic.¹⁴ In both phases of assessment, Quantra correctly classified these categories, which form the midpoint between high and low densities, with a few cases of BI-RADS 2 graded as 3 and *vice versa* (Figure 1a,b).

Although there was an excellent distinction between low-grade and high-grade densities, Quantra graded most of the BI-RADS 1 categories as BI-RADS 2 in both phases (Figure 1a,b). Restricting the analysis to Phase 2 majority report, we observed a 10.3% increase from 25.4% to 35.7% in sensitivity and a 4% decrease in specificity for BI-RADS 1 (Table 5). The result also shows that Quantra

classified 42% and 47% of BI-RADS 4 as 3 in Phases 1 and 2, respectively. It is possible that the discrepancy between BI-RADS and Quantra for BI-RADS 3 and 4 may be due to radiologist assessment being influenced by perceived masking effect (concealment of cancer due to dense tissue superimposition), which usually warrants further imaging,¹⁴ a phenomenon not considered by Quantra. A further observation is that a small difference in Quantra's total fractional quantized breast density (q_{abd}) values contributed to the discrepancy between BI-RADS and Quantra rating. This effect was most pronounced for BI-RADS 3 and 4 (Figure 2d) than between BI-RADS 2 and 3 (Figure 2c). Therefore, it can be argued that Quantra has a narrow threshold for BI-RADS 1 and a wider threshold for BI-RADS 3 and may need to be recalibrated to improve the sensitivity for BI-RADS 1 and 4.

Whilst it is desirable for Quantra to absolutely reproduce BI-RADS, it should be remembered that Quantra measures volumetric breast density which may be more related to the quantity of the dense tissue that is the major determinant breast cancer risk.^{32–34} The

Table 3. Agreement between Quantra™ (Hologic Inc., Bedford, MA) and Breast Imaging Reporting and Data Systems (BI-RADS®) (95% confidence interval) in Phase 2

Quantra vs BI-RADS	Simple kappa four-grade scale	Weighted kappa four-grade scale	Simple kappa two-grade scale	Weighted kappa two-grade scale
Q vs R1	0.47 (0.41–0.53)	0.77 (0.72–0.81)	0.50 (0.42–0.57)	0.81 (0.74–0.86)
Q vs R2	0.55 (0.48–0.62)	0.81 (0.76–0.84)	0.57 (0.50–0.64)	0.83 (0.79–0.87)
Q vs R3	0.48 (0.41–0.55)	0.78 (0.74–0.85)	0.50 (0.43–0.57)	0.81 (0.77–0.85)
Q vs MR	0.55 (0.48–0.63)	0.79 (0.75–0.84)	0.57 (0.50–0.64)	0.84 (0.79–0.87)

MR, majority report; Q, Quantra; R, reader.

Simple kappa: percentage agreement.

Weighted kappa: agreement beyond chance expectation.

Table 4. Performance of Quantra™ (Hologic Inc., Bedford, MA) in reproducing Breast Imaging Reporting and Data Systems (BI-RADS®) assessment in Phase 1; it summarizes the performance of Quantra in reproducing BI-RADS as measured by the area under curve (AUC) with sensitivity and specificity

BI-RADS	AUC	SE	95% CI	Sensitivity %	Specificity %
BI-RADS 1	0.861	0.012	0.84–0.88	25.4	98.9
BI-RADS 2	0.784	0.013	0.76–0.81	81.1	75.4
BI-RADS 3	0.707	0.014	0.68–0.74	90.7	61.3
BI-RADS 4	0.862	0.012	0.84–0.89	56.7	95.3
BI-RADS 1–2 vs 3–4	0.917	0.008	0.90–0.93	93.2	86.1

BI-RADS 1–2, almost fatty to scattered fibroglandular tissue; BI-RADS 3–4, heterogeneously dense to extremely dense breasts; CI, confidence interval; SE, standard error.

association between breast density and cancer risk emphasizes the need to consider dense tissue volume when estimating breast density, particularly for the purpose of risk assessment. This is because even with the same size of the projected dense area, the thickness and volume of the dense tissue may not necessarily be the same in different females³⁵ and is perhaps the reason why volumetric MBD assessment methods are better risk predictors from MBD calculations than visual and area-based approaches.¹⁶

The result of this study is comparable with that of previous studies assessing the performance of Quantra in reproducing BI-RADS. Unlike previous studies that focused on the differentiation of high- and low-grade densities, the present work has also shown the performance of Quantra on a four-grade scale (reproducibility of each BI-RADS category). An important observation in the present study is that Quantra is a poor predictor of BI-RADS on a four-grade scale, but a small difference in the total fractional quantized breast density (q_abd) value may be a significant cause of this reduced reproducibility. Therefore, Quantra should be considered a tool for two-grade scale MBD classification.

Previous studies have reported sensitivity and specificity on a two-grade scale ranging from 78.0% to 89.8% and 84.6–88.9%, respectively, for Quantra. The reported cut-off values for Quantra in differentiating high-grade from low-grade BI-RADS categories ranged from 13% to 22%.^{23–25,36} The variation in sensitivity, specificity and cut-off values in the different studies may be due to the varying sample sizes, test sets and radiologists. For example, considering the majority report in Phase 2, a cut-off value of 20.5% was observed. Even though Quantra studies have shown variable cut-off values, the differences in the majority of the

studies are negligible (19.5–22%);^{23,24,36} only one study has reported a lower cut-off value (13%).²⁵ Ciatto *et al*²³ have questioned this lower cut-off value, implicating differences in the radiologist panels as the cause for the lower value. Furthermore, the 13% cut-off reported by Rafferty *et al*²⁵ has not yet been peer reviewed.

The first thorough independent evaluation of Quantra was performed by Ciatto *et al*²³ using a test set of 481 mammograms and 11 radiologists and reported correct classification of 88.6% of D1–2 (almost fatty to scattered fibroglandular tissue) and 89.8% of D3–4 (heterogeneously dense to extremely dense breasts) BI-RADS categories by Quantra. The authors assessed the performance of Quantra in reproducing BI-RADS assessment on a two-category scale, neglecting the four-category scale used by radiologists clinically. In addition, their study was based on the first version of Quantra (v. 1.0) and on a test-set reading, which may differ from the actual clinical scenario. Technology is constantly evolving and newer versions of Quantra have been developed, including recalibration and upgrade in technology to improve performance. Without further studies, it is impossible to establish whether a new version of Quantra is fit for the purpose of MBD assessment or whether it demonstrates improvement over older versions. The present study differs from the only peer-reviewed studies by Ciatto *et al*²³ and Regini *et al*²⁴ in that it is based on a new version of Quantra (v. 2.0) and evaluated its performance in reproducing BI-RADS rating on a four-grade scale commonly used in the clinical setting. It also provides evidence for the performance of Quantra in reproducing BI-RADS categories assigned clinically and in a test-set reading. Thus, the present study has provided

Table 5. Performance of Quantra™ (Hologic Inc., Bedford, MA) in reproducing Breast Imaging Reporting and Data Systems (BI-RADS®) categories assigned in Phase 2 (majority report)

BI-RADS	AUC	SE	95% CI	Sensitivity %	Specificity %
BI-RADS 1	0.823	0.033	0.759–0.886	35.7	94.9
BI-RADS 2	0.822	0.025	0.772–0.871	91.2	77.7
BI-RADS 3	0.737	0.030	0.626–0.745	88.6	61.0
BI-RADS 4	0.829	0.026	0.779–0.879	50.3	96.7
BI-RADS1–2 vs 3–4	0.885	0.023	0.815–0.907	91.3	83.6

AUC, area under the curve; BI-RADS 1–2, almost fatty to scattered fibroglandular tissue; BI-RADS 3–4, heterogeneously dense to extremely dense breasts; CI, confidence interval; SE, standard error.

Figure 2. Examples of mammograms showing types of cases where radiologists and Quantra™ (Hologic Inc., Bedford, MA) more consistently showed similarity (a, b) and where they more consistently showed discrepancy in grading (c, d). In (c) the radiologists assigned Grade 2 and Quantra assigned the density Grade 3. In (d) the radiologists assigned Grade 4 and Quantra assigned Grade 3. Quantra total quantized breast density for the cases was: (a) 3.47 (b) 3.36, (c) 2.51 and (d) 3.43.

a new insight into the clinical performance of a newer version of Quantra (v2.0), highlighting its limitation on a four-grade scale and improved performance on a two-grade scale. Quantra has been shown to be a reproducible tool for volumetric breast density assessment.^{37–39} Findings suggest that Quantra may be a useful supplementary tool to BI-RADS for MBD assessment. Automated MBD assessment may also eliminate variability in MBD classification, meaning a more consistent use of breast density information in clinical decision-making.

Limitations of this study include that six radiologists from the same practice were involved across the BI-RADS breast density reports in Phase 1. Variability in MBD assessment has been reported in the literature with interreader agreement (κ) ranging from 0.19 to 0.91.^{15,18,19} This interreader variability is also evident in Phase 2 of the present study (Table 2). The high level of BI-RADS interreader agreement observed in the present study may be due to all readers working in the same practice. It is possible that they would demonstrate considerable interreader variability with readers from different practice, limiting the generalizability of results. Using the majority report of radiologists to assess the performance of Quantra is a better reference standard. Therefore, if we were to use the majority report of the six radiologists in Phase 1, some of the BI-RADS categories may have been different. However, we have also assessed the performance of Quantra using the majority report in Phase 2. It is also possible that the increased

sensitivity of Quantra for BI-RADS 1 and 2 in Phase 2 may be due to the small sample size compared with sample in Phase 1 and the laboratory effect.²⁷

Conversely, the strength of the study is that BI-RADS categories were assigned by breast radiologists using BI-RADS scheme on a daily basis. Also, the high level of intrareader and interreader agreement, the agreement between each radiologist and Quantra and the agreement between the majority report in Phase 2 and the report in Phase 1 show that the latter can be relied on to assess the performance of Quantra in reproducing BI-RADS categories. A further strength of this study is the large sample size compared with previous studies. Also, since extraneous variables influence radiologist behaviour in a laboratory environment,²⁶ the fact that we assessed the performance of Quantra in reproducing BI-RADS with categories assigned clinically and in a laboratory setting adds to the strengths of this work.

CONCLUSION

The result demonstrates that Quantra is very good at reproducing BI-RADS classification on a two-grade scale. A cut-off value of 20% allowed better differentiation between high-grade and low-grade categories using Quantra. Although discrepancies are evident between BI-RADS 1 and 2, and between BI-RADS 3 and 4, Quantra is a promising adjunct to BI-RADS for MBD assessment in clinical practice.

REFERENCES

1. 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**: 227–36. doi: <http://dx.doi.org/10.1056/NEJMoa062790>
2. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. *Radiology* 2008; **246**: 376–83. doi: <http://dx.doi.org/10.1148/radiol.2461070200>
3. Njor S, Nystrom L, Moss S, Paci E, Broeders M, Segnan N, et al. Breast cancer mortality in mammographic screening in Europe:

- a review of incidence-based mortality studies. *J Med Screen* 2012; **19** (Suppl. 1): 33–41. doi: <http://dx.doi.org/10.1258/jms.2012.012080>
4. Paci E; EUROSREEN Working Group. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* 2012; **19**(Suppl. 1): 5–13. doi: <http://dx.doi.org/10.1258/jms.2012.012077>
 5. Ekpo EU, Hogg P, Highnam R, McEntee MF. Breast composition: measurement and clinical use. *Radiography* 2015; **21**: 324–33. doi: <http://dx.doi.org/10.1016/j.radi.2015.06.006>
 6. Cabanes A, Pastor-Barriuso R, García-López M, Pedraz-Pingarrón C, Sánchez-Contador C, Vázquez Carrete JA, et al. Alcohol, tobacco, and mammographic density: a population-based study. *Breast Cancer Res Treat* 2011; **129**: 135–47. doi: <http://dx.doi.org/10.1007/s10549-011-1414-5>
 7. Yaghjian L, Mahoney MC, Succop P, Wones R, Buckholz J, Pinney SM. Relationship between breast cancer risk factors and mammographic breast density in the Fernald Community Cohort. *Br J Cancer* 2012; **106**: 996–1003. doi: <http://dx.doi.org/10.1038/bjc.2012.1>
 8. Cuzick J, Warwick J, Pinney E, Duffy SW, Cawthorn S, Howell A, et al. Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: a nested case-control study. *J Natl Cancer Inst* 2011; **103**: 744–52. doi: <http://dx.doi.org/10.1093/jnci/djr079>
 9. Li J, Humphreys K, Eriksson L, Edgren G, Czene K, Hall P. Mammographic density reduction is a prognostic marker of response to adjuvant tamoxifen therapy in postmenopausal patients with breast cancer. *J Clin Oncol* 2013; **31**: 2249–56. doi: <http://dx.doi.org/10.1200/JCO.2012.44.5015>
 10. Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst* 2006; **98**: 1204–14. doi: <http://dx.doi.org/10.1093/jnci/djj331>
 11. Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Böhm-Vélez M, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* 2008; **299**: 2151–63. doi: <http://dx.doi.org/10.1001/jama.299.18.2151>
 12. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. 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* 2012; **307**: 1394–404. doi: <http://dx.doi.org/10.1001/jama.2012.388>
 13. Rafferty EA, Park JM, Philpotts LE, Poplack SP, Sumkin JH, Halpern EF, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology* 2013; **266**: 104–13. doi: <http://dx.doi.org/10.1148/radiol.12120674>
 14. Pinsky RW, Helvie MA. Mammographic breast density: effect on imaging and breast cancer risk. *J Natl Compr Canc Netw* 2010; **8**: 1157–64.
 15. Ekpo EU, McEntee MF. Measurement of breast density with digital breast tomosynthesis—a systematic review. *Br J Radiol* 2014; **87**: 20140460. doi: <http://dx.doi.org/10.1259/bjr.20140460>
 16. Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, et al. Digital mammographic density and breast cancer risk: a case inverted question mark control study of six alternative density assessment methods. *Breast Cancer Res* 2014; **16**: 439. doi: <http://dx.doi.org/10.1186/s13058-014-0439-1>
 17. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 1159–69. doi: <http://dx.doi.org/10.1158/1055-9965.EPI-06-0034>
 18. Bernardi D, Pellegrini M, Di Michele S, Tuttobene P, Fantó C, Valentini M, et al. Interobserver agreement in breast radiological density attribution according to BI-RADS quantitative classification. *Radiol Med* 2012; **117**: 519–28. doi: <http://dx.doi.org/10.1007/s11547-011-0777-3>
 19. Spayne MC, Gard CC, Skelly J, Miglioretti DL, Vacek PM, Geller BM. Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study. *Breast J* 2012; **18**: 326–33. doi: <http://dx.doi.org/10.1111/j.1524-4741.2012.01250.x>
 20. Diffey J, Berks M, Hufton A, Chung C, Verow R, Morrison J, et al. A stepwedge-based method for measuring breast density: observer variability and comparison with human reading. In: Samei E, Pelc NJ, eds. *Medical imaging 2010: physics of medical imaging*. Bellingham, WA: Spie-Int Soc Optical Engineering; 2010.
 21. Kopans DB. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. *Radiology* 2008; **246**: 348–53. doi: <http://dx.doi.org/10.1148/radiol.2461070309>
 22. van der Waal D, den Heeten GJ, Pijnappel RM, Schuur KH, Timmers JM, Verbeek AL, et al. 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* 2015; **10**: e0136667. doi: <http://dx.doi.org/10.1371/journal.pone.0136667>
 23. Ciatto S, Bernardi D, Calabrese M, Durando M, Gentilini MA, Mariscotti G, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. *Breast* 2012; **21**: 503–6. doi: <http://dx.doi.org/10.1016/j.breast.2012.01.005>
 24. Regini E, Mariscotti G, Durando M, Ghione G, Luparia A, Campanino PP, et al. Radiological assessment of breast density by visual classification (BI-RADS) compared to automated volumetric digital software (Quantra): implications for clinical practice. *Radiol Med* 2014; **119**: 741–9. doi: <http://dx.doi.org/10.1007/s11547-014-0390-3>
 25. Rafferty E, Smith A, Niklason L. *Comparison of three methods of estimating breast density: BI-RADS density scores using full field digital mammography, breast tomosynthesis, and volumetric breast density*. Chicago, IL: Radiological Society of North America; 2009.
 26. Soh BP, Lee W, Kench PL, Reed WM, McEntee MF, Poulos A, et al. Assessing reader performance in radiology, an imperfect science: lessons from breast screening. *Clin Radiol* 2012; **67**: 623–8. doi: <http://dx.doi.org/10.1016/j.crad.2012.02.007>
 27. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The “laboratory” effect: comparing radiologists’ performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology* 2008; **249**: 47–53. doi: <http://dx.doi.org/10.1148/radiol.2491072025>
 28. Highnam R, Brady M, Yaffe MJ, Karssemeijer N, Harvey J. Robust breast composition measurement—Volpara (TM). In: Marti J, Oliver A, Freixenet J, Marti R, eds. *Digital mammography*. Berlin, Germany: Springer-Verlag Berlin; 2010. pp. 342–9.
 29. Ren B, Smith A, Jing Z. *Measurement of breast density with digital breast tomosynthesis*. San Diego, CA: Physics of Medical Imaging: The Society of Photo-optical Instrumentation Engineers; 2012.
 30. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005; **37**: 360–3.
 31. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves: pedagogical tools and methods. *CJEM* 2006; **8**: 19–20.

32. Boyd NF, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst* 2010; **102**: 1224–37. doi: <http://dx.doi.org/10.1093/jnci/djq239>
33. Alowami S, Troup S, Al-Haddad S, Kirkpatrick I, Watson PH. Mammographic density is related to stroma and stromal proteoglycan expression. *Breast Cancer Res* 2003; **5**: R129–35. doi: <http://dx.doi.org/10.1186/bcr622>
34. Hawes D, Downey S, Pearce CL, Bartow S, Wan P, Pike MC, et al. Dense breast stromal tissue shows greatly increased concentration of breast epithelium but no increase in its proliferative activity. *Breast Cancer Res* 2006; **8**: R24. doi: <http://dx.doi.org/10.1186/bcr1408>
35. Boyd N, Martin L, Gunasekar A, Melnichouk O, Maudsley G, Peressotti C, et al. Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 1754–62. doi: <http://dx.doi.org/10.1158/1055-9965.EPI-09-0107>
36. Pahwa S, Hari S, Thulkar S, Angraal S. Evaluation of breast parenchymal density with QUANTRA software. Vienna, Austria; ECR, 6–10 March 2014; Vienna, Austria. European Congress of Radiology, 2014.
37. Singh JM, Fallenberg EM, Diekmann F, Renz DM, Witlandt R, Bick U, et al. Volumetric Breast Density Assessment: Reproducibility in Serial Examinations and Comparison with Visual Assessment. Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren; 2013.
38. Engelken F, Singh JM, Fallenberg EM, Bick U, Böttcher J, Renz DM. Volumetric breast composition analysis: reproducibility of breast percent density and fibroglandular tissue volume measurements in serial mammograms. *Acta Radiol* 2014; **55**: 32–8. doi: <http://dx.doi.org/10.1177/0284185113492721>
39. Schmachtenberg C, Hammann-Kloss S, Bick U, Engelken F. Intraindividual comparison of two methods of volumetric breast composition assessment. *Acad Radiol* 2015; **22**: 447–52. doi: <http://dx.doi.org/10.1016/j.acra.2014.12.003>