WHITE PAPER

TruPGMI: AI for mammography quality improvement

2022

Lisa Johnston, PhD
Lead Product Manager
Volpara Health
lisa.johnston@volparahealth.com

Ralph Highnam, PhD
Chief Science and Innovation Officer
Volpara Health
ralph.highnam@volparahealth.com

TruPGMI: AI for mammography quality improvement

Mammography is a common breast screening practice across the world. However, the standards used to gauge the quality of mammography have varied between countries and regions.

Introduction

This paper focuses on the patient positioning aspects of image quality in mammography, with a review of prior practices and an explanation of the process that led to development of the Volpara® TruPGMI™ algorithm, which uses artificial intelligence (AI) to automatically evaluate the image quality of every standard mammographic view. It outlines how TruPGMI addresses issues with historic image evaluation systems, and how TruPGMI can be used to target technologist training to improve the quality of mammographic positioning.

Despite widespread respect for the work of Bassett,[1] Eklund,[2] and others, and even though there has been some harmonization in countries with population-based screening, in general the quality of mammographic positioning has not been standardized.

The impact of positioning

Mammographic positioning affects clinical performance in numerous ways that are well documented in the literature.[3,4,5] Positioning skill affects the sensitivity and specificity of mammography, and poor positioning is widely believed to be a significant cause of technical recalls. The subjective and variable nature of observer assessment of positioning also reduces the effectiveness of feedback to the technologist.

Sensitivity of mammography. Like Buist,[4] Taplin[5] showed that the sensitivity of mammography fell from 84% to 66.3% when image positioning quality dropped from **meeting positioning criteria** to **failing to meet positioning criteria.** This has remained an issue for many years and has led to efforts such as the FDA Enhancing Quality Using the Inspection Program (EQUIP) initiative.

Specificity of mammography. Poor positioning can result in overlapping glandular tissue, causing both false negatives (masking) and false positives (masquerading), which directly affects the specificity of mammography. While maintaining high sensitivity in mammography is critical, it is the specificity that impacts financial cost, patient anxiety, and screening compliance.[6]

Reduction of technical recalls. Poor positioning often results in technical recalls that are expensive, create scheduling issues, and cause anxiety and inconvenience to the patient.

Subjectivity and observer variability. Traditional assessment of patient positioning is subjective[7] and prone to observer variability, as this paper will show. In fact, observers cannot even agree on how to rank the importance of the characteristics used to assess patient positioning.

Improving positioning with consistent and objective feedback

Having a truly objective means of assessing mammographic positioning would help identify areas for performance improvement. Objective feedback offers each technologist quality assessment that is consistent, fair, and measurable.

The more often feedback is provided, the better a technologist can gauge true performance and analyze what improvements could be made. Ideally, the technologist is sufficiently trained such that the number of technical recalls is reduced.

Prior image evaluation systems

In 1994, the UK National Health System Breast Screening Programme published the PGMI Image Evaluation System,[8,9] a method of visually assessing mammograms. The premise was to evaluate positioning and other features and to use those features to classify studies as **Perfect, Good, Moderate, or Inadequate (P, G, M, or I).**

The only practical way to utilize such a visual assessment method was sampling. The UK chose to sample 50 mammograms acquired by each screening technologist, every three years. The results of the assessment were provided to the technologists as a means of encouraging them to improve.

Visual PGMI variants

PGMI-like systems are used in a number of countries (table 1), but they differ considerably from one another. For instance, the original UK PGMI system used 16 different metrics, while the Norwegian system uses 38.

Visual PGMI assessment of positioning

Several factors contribute to the subjective nature of visual assessment of mammography positioning.

Lack of consensus on evaluation criteria. The work of Spuur[10] demonstrates "a lack of consensus among technologists in the interpretation of criteria in current image evaluation systems in mammography"; governing bodies, radiologists, and technologists cannot agree on the criteria for good positioning. When comparing the criteria included in guidelines across various countries, the substantial differences are clear (table 2).

Inter-observer variability. The lack of consensus on evaluation criteria may increase inter-observer variability. Boyce[11] evaluated inter-observer variability among technologists from the UK and Norway who are trained to use different variants of PGMI. The results showed a great deal of inter-observer variability, with a slight preference for images from their own country.

Taylor[12] found that inter-observer agreement across an expert panel was somewhat limited, even after the panel first came to a consensus on the criteria and specific wording for the image quality assessment.

Intra-observer variability. The human practice of assigning things to categories is variable by its very nature,[13,14] as shown so clearly in studies of radiologists assigning BI-RADS breast density categories.[15]

Taylor acknowledges in her work that "A further limitation includes some inherent subjectivity associated with this type of image assessment" but adds "the large numbers of images for analysis help minimize intra observer variability."[12] In screening, a technologist must assess if an image is well positioned but does not have the benefit of the feedback on "large numbers of images" and may not receive regular feedback.

Table 1. National guidelines for image quality

Country	Guideline
Australia	PGMI adapted from UK version
Denmark	PGMI (adopted 1995)
New Zealand	PGMI adapted from UK version
Norway	Norwegian Breast Cancer Screening Program (NBCSP) criteria expanded from PGMI
UK	PGMI standard

Table 2. Comparison of positioning criteria included in guidelines across various countries

Country		United States	United Kingdom	Australia	Netherlands	Belgium
Guidelines maintained by		American College of Radiology (23)	National Health Service (13)	BreastScreen Australia (14)	LRCB, Dutch Expert Centre for Screening (24)	European Commission (25)
Criteria (CC/MLO views)	Skin folds	✔	✔	✔	✔	✔
	Asymmetry left vs. right		✔	✔	✔	✔
	Nipple is not in profile		✔	✔	✔	✔
	PNL>1cm between CC and MLO views	✔		✔		
	Breast tissue cutoff	✔		✔		✔
	Absence of artifacts/other body parts	✔	✔	✔	✔	✔
	Fibroglandular disc/ triangle				✔	
Criteria (CC view only)	Medial tissue not visualized		✔	✔	✔	✔
	Lateral tissue/axillary tail not visualized		✔		✔	✔
	Posterior tissue not visualized	✔	✔			✔
	Pectoral muscle/shadown visualized		✔		✔	✔
	Excessive exaggeration	✔				
Criteria (MLO view only)	IMF not well demonstrated/visualized	✔	✔	✔	✔	✔
	IMF skin folds of IMF obscured					✔
	Pectoral length to level of nipple/PNL		✔			✔
	Pectoral angle		✔			✔
	Full width/sufficient amount of pectoral muscle			✔	✔	
	Breast too high on receptor	✔				
	Breast sag/droop	✔	✔			✔
	Posterior tissue not visualized	✔	✔	✔		

In the UK: visual PGMI without the PGMI score

The UK recently has moved away from the practice of assigning the PGMI category score to mammograms because of its subjectivity. Instead, technologists are now judged simply on the individual metrics that comprise PGMI.

While the UK approach is understandable, PGMI scoring has several important benefits:

- Sorting by quality when searching for high-quality studies for accreditation
- Sorting by quality when searching for low-quality studies for training and assigning corrective actions
- Understanding overall performance trends
- Comparing positioning performance among individuals, between groups of technologists, or between regions or nations

Sampling is not sufficient

The original UK approach of sampling 50 cases per technologist every three years is simply not sufficient as a means of monitoring quality. A technologist's poor technique, for example, might go undetected for as long as three years.

Because visual assessment is very time-consuming and subjective, sampling was the only reasonable approach to PGMI; that is, until automation became available.

Automation of PGMI

In theory, an objective, automated version of PGMI would reduce observer variability, provide technologists with regular performance feedback, reduce technical recalls, and improve clinical performance.

Intra- and inter-observer variability can be completely addressed through automation, where an algorithm is used to assess the images, rather than a human. The algorithm, unlike human observers, will always score the same images the same way.

Automation is critical for standardization and workflow. The assessment of patient positioning provides a primary source of feedback to the technologist. Automated evaluation enables frequent feedback which is much more likely to drive continuous quality improvement compared with sporadic feedback based on an anecdotal visual assessment.

The TruPGMI algorithm

Volpara Health created the TruPGMI algorithm with the aim of helping technologists further develop their positioning skills and receive relevant on-the-job training. Now in its third generation, TruPGMI uses AI to automatically evaluate the image quality of every standard mammographic view at a breast imaging facility, identifying any positioning deficiencies and then assigning the images a score of P, G, M, or I.

TruPGMI is designed specifically for automation. It incorporates definitions of metrics that can be measured precisely time and time again; TruPGMI does not, for instance, include a metric for "sufficient pectoral muscle" because "sufficient" is not measurable.

Manual assessment of image quality is time consuming, which is why many image evaluation systems suggest randomly sampling a small number of images. TruPGMI's use of AI means that all images can be assessed for positioning quality, allowing evaluation of every mammogram, rather than an image sample, providing a more accurate gauge of image quality.

In total, TruPGMI has been applied to over 60 million images taken by over 5,000 technologists across North America, New Zealand, Australia, and Europe.

Derivation of the TruPGMI standard

The metrics underlying TruPGMI were derived using best practices from around the world, including those used both in population-based screening programs and opportunistic screening, and were developed over time by reader studies and clinical experts. The metrics chosen to be included in TruPGMI needed to be measurable criteria so that AI can be used to accurately segment structures and obtain consistent measurements.

The metrics included in TruPGMI

- directly relate to features that technologists can observe in images; and
- are under the control of the technologist.

The diagram below shows the metrics assessed by TruPGMI (figure 1).

Figure 1. TruPGMI metrics

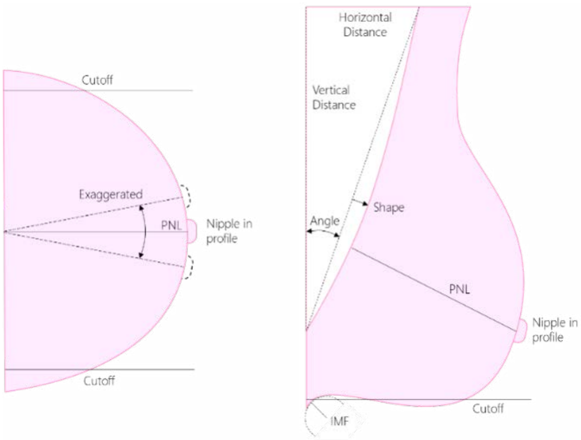

Table 3. Percentage of images that meet each TruPGMI metric taken by technologists in the 50th and 90th percentiles

Positioning Metrics	Global Median	Top 10%
CC nipple in profile	79%	88%
CC PNL met	68%	80%
CC nipple midline	46%	56%
CC no cutoff	98%	100%
MLO nipple in profile	82%	90%
MLO adequate pec	93%	97%
MLO IMF visible	36%	52%
MLO pec no skin folds	96%	99%
MLO pec to PNL met	65%	79%
MLO no cutoff	99%	100%
MLO no concave pec	76%	87%

TruPGMI category assignment

TruPGMI first evaluates each image individually, giving it a numeric score based on how well it meets each required criterion (presence, absence, or degree of a feature's presence). A numeric study score is then compiled from the numeric images scores. Each image and the study are then assigned to the P, G, M, or I category.

TruPGMI in practice

TruPGMI is not only useful in helping screening facilities to fulfill their regulatory and compliance requirements by simplifying the image selection process for submission, but is also a key tool for benchmarking.

Benchmarking in mammography using TruPGMI

As in any business, benchmarking is important in mammography to know what best practices look like. Benchmarks can be used to understand how individuals perform and what is achievable in clinical practice. Key image quality metrics can be measured and compared within an imaging facility, against peer facilities nationwide, or even globally. Understanding one's performance relative to benchmarks helps technologists and managers set achievable, realistic goals for performance improvement.

The larger the dataset used to create the benchmarks, the more accurate and useful they will be. Outlying data points will have less impact, allowing trends to be clearly identified. Previous studies reporting on benchmarking statistics in mammography have been limited to small datasets due to the time-consuming nature of manual image evaluation.[1,7] Volpara has the largest known cohort of images from a global database. A subset of 3.3 million images from 2000 technologists has been analyzed to formulate key benchmarking statistics based on the current clinical climate (table 3).

Conclusion

TruPGMI is an objective, automated scoring system that eliminates the impact of inter- and intra-reader variability. It provides continuous, fair, and meaningful feedback that helps managers provide directed training to their staff and helps technologists advance their positioning techniques.

TruPGMI contains a clear, defined set of metrics that evaluate patient positioning in a manner that supports complete automation. TruPGMI is designed to be international in scope and will lead to better imaging wherever it is applied.

Volpara TruPGMI is available as a clinical function within Volpara® Analytics™ software as a component of the Volpara® Breast Health Platform™.

Because TruPGMI is completely automated, it lends itself to use on every mammogram, rather than just small samples. To date, TruPGMI has been applied to more than 60 million mammograms (both 2D and 3D) from North America, New Zealand, Australia, and Europe, representing the largest known automated breast positioning assessment cohort.

1. Bassett LW et al., Mammographic positioning: evaluation from the view box, Radiology. 1993 Sep;188(3):803–6.

2. Eklund GW et al., Assessing adequacy of mammographic image quality, Radiology. 1994 Feb;190(2):297–307.

3. Bae MS et al., Breast Cancer Detected with Screening US: Reasons for Nondetection at Mammography. Radiology: Volume 270: Number 2—February 2014.

4. Buist DSM et al., Factors contributing to mammography failure in women aged 40–49 years, Journal of the National Cancer Institute 96(19), 1432–1440 (2004).

5. Taplin SH et al., Screening Mammography: Clinical Image Quality and the Risk of Interval Breast CancerAJR:178, April 2002.

6. Drukteinis JS, Mooney BP, Flowers CI, Gatenby RA. Beyond mammography: new frontiers in breast cancer screening. Am J Med. 2013;126(6):472–479. doi:10.1016/j.amjmed.2012.11.025

7. Huppe AI, Overman KL, Gatewood JB, Hill JD, Miller LC, Inciardi MF. Mammography positioning standards in the digital era: is the status quo acceptable? AJR American journal of roentgenology. 2017;209 (6):1419-25.

8. Guidelines for Quality Assurance in Mammography Screening—THIRD EDITION, Published by The National Cancer Screening Service Board, King's Inns House, 200 Parnell Street, Dublin 1, Ireland.

9. NHS Breast Screening Programme. Quality assurance guidelines for technologists. 30. Sheffield: NHS Breast Screening Programme, 1994.

10. Spuur K et al., Evaluation of the pectoral muscle in mammography images: The Australian experience, European Journal of Radiography (2009) 1, 12–21.

11. Boyce M et al., Comparing the use and interpretation of PGMI scoring to assess the technical quality of screening mammograms in the UK and Norway, Radiography 21 (2015) 342–347.

12. Taylor K et al., Mammographic image quality in relation to positioning of the breast: A multicentre international evaluation of the assessment systems currently used, to provide an evidence base for establishing a standardised method of assessment, Radiography 23 (2017) 343–349.

13. Eom HJ et al., Comparison of variability in breast density assessment by BI-RADS category according to the level of experience, Acta Radiol. 2018 May;59(5):527–532.

14. Mitani AA et al., Summary measures of agreement and association between many raters' ordinal classifications, Ann Epidemiol. 2017 Oct;27(10):677–685.

15. Sprague BL et al., Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice, Ann Intern Med. 2016 Oct 4;165(7):457–464.

Contact

info@volparahealth.com
support@volparahealth.com
volparahealth.com

US +1 855 607 0478
AUS 1800 370 623
NZ 0800 444 148

Europe +44 203 051 1029
Global +64 4 499 6029

Connect

@VolparaHealth
@volpara
Volpara Health

Copyright © 2022 Volpara Health Technologies Ltd. All rights reserved.