Radiography xxx (2017) 1–7

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

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

K. Taylor [a, *], D. Parashar [b, c], G. Bouverat [d], A. Poulos [e], R. Gullien [f], E. Stewart [g], R. Aarre [h], P. Crystal [i], M. Wallis [a]

[a] Cambridge Breast Unit, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ, England, United Kingdom
[b] Cancer Research Unit, Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry, CV4 7AL, England, United Kingdom
[c] Cancer Research Centre, University of Warwick, Coventry, CV4 7AL, England, United Kingdom
[d] Statistics and Epidemiology Unit, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, CV4 7AL, England, United Kingdom
[e] Faculty of Health Sciences, The University of Sydney, Camperdown, NSW, 2006, Australia
[f] Breast Imaging Centre, Oslo University Hospital, Oslo, Norway
[g] Breast Screen Victoria, 78 Gheringhap Street, Geelong, 3220, VIC, Australia
[h] Hvidovre Hospital, Mammografiklinikken, Afsnit 034, Pavillon 1, Østvej, 2650, Hvidovre, Denmark
[i] Mount Sinai Hospital, Women's College Hospital, Room 574, Mount Sinai Hospital, 600 University Avenue, Toronto, ON M5G 1X5, Canada

ARTICLE INFO

Article history:
Received 25 November 2016
Received in revised form
13 February 2017
Accepted 5 March 2017
Available online xxx

Keywords:
Image quality
Mammograms
Weighted criteria
Validation

ABSTRACT

Introduction: Optimum mammography positioning technique is necessary to maximise cancer detection. Current criteria for mammography appraisal lack reliability and validity with a need to develop a more objective system.
We aimed to establish current international practice in assessing image quality (IQ), of screening mammograms then develop and validate a reproducible assessment tool.
Methods: A questionnaire sent to centres in countries undertaking population screening identified practice, participants for an expert panel (EP) of radiologists/radiographers and a testing panel (TP) of radiographers. The EP developed category criteria and descriptors using a modified Delphi process to agree definitions.
The EP scored 12 screening mammograms to test agreement then a main set of 178 cases. Weighted scores were derived for each descriptor enabling calculation of numerical parameters for each new category. The TP then scored the main set. Statistical analysis included ANOVA, t-tests and Kendall's coefficient.
Results: 11 centres in 8 countries responded forming an EP of 7 members and TP of 44 members.
The EP showed moderate agreement when the scoring the mini test set $W = 0.50$ $p < 0.001$ and the main set $W = 0.55$ $p < 0.001$, 'posterior nipple line' being the most difficult descriptor.
The weighted total scores differentiated the 4 new categories Perfect, Good, Adequate and Inadequate ($p < 0.001$).
Conclusion: We have developed an assessment tool by Delphi consensus and weighted consensus criteria. We have successfully tabulated a range of numerical scores for each new category providing the first validated and reproducible mammography IQ scoring system.

Introduction

In order to achieve a high quality diagnostic mammogram, a number of factors need to be considered, not least the expertise of the mammographer in producing optimally positioned breasts

* Corresponding author.
E-mail addresses: kathryn.taylor@addenbrookes.nhs.uk (K. Taylor), D.parashar@warwick.ac.uk (D. Parashar), g.bouverat@warwick.ac.uk (G. Bouverat), ann.poulos@sydney.edu.au (A. Poulos), uxraul@ous-hf.no (R. Gullien), estewart@breastscreen.org.au (E. Stewart), Randi.Aarre@regionh.dk (R. Aarre), Pavel.crystal@sinaihealthsystem.ca (P. Crystal), Matthew.wallis@addenbrookes.nhs.uk (M. Wallis).

http://dx.doi.org/10.1016/j.radi.2017.03.004
1078-8174/

(Fig 1). Positioning has been cited as the single most important factor in optimising mammographic image quality (IQ). Without all breast tissue included on a mammogram or not optimally visualised, all other aspects of IQ are not relevant. Research has shown a direct link between mammographic image quality IQ and cancer detection.[1] Optimal IQ leads to earlier detection, higher detection rates, fewer interval cancers and reduced dose.[2–4]

Current UK categories for mammography evaluation are PGMI (Perfect, Good, Moderate and Inadequate). Use of PGMI was established by the National Health Service breast screening programme in 2006, and there is evidence this has been adopted by other countries.[5–7] Some parts of Europe and the United States use evaluation tools provided by the Commission of the European Communities (CEC) and the American College of Radiologists (ACR) respectively.[8–10] The common theme is a list of categories and associated criteria which largely relate to positioning of the breast. These are used to determine IQ that informs the assignment of the image as excellent, acceptable or inadequate quality. There is evidence that these systems currently lack reliability and validity; guidelines for their implementation have always been subjective and have also not been reviewed commensurate with altered imaging practice such as the move from analogue to digital image acquisition.[11,12]

Difficulties involved in developing and validating any image assessment tool are twofold. First deciding which anatomical structures should be included in the image then the level of importance given by the observer to the inclusion of each structure. There is a documented need to develop a visual grading scale for consistency in evaluating image quality[7,13] combining both aspects of the assessment.

In this study we will first establish current international practice in assessing the IQ of screening mammograms. Then develop and validate an assessment tool, incorporating a weighted consensus list of criteria derived from currents systems and deemed most relevant when assessing mammographic IQ.

Methods

The study comprised several phases. Participants completed an on line questionnaire to establish current practice in assessing the IQ of screening mammograms. Using PGMI as a starting point, a sub group (expert panel) employed a Delphi process and test set of mammograms to develop and test revised, weighted criteria and numerical ranges for overall scoring categories.

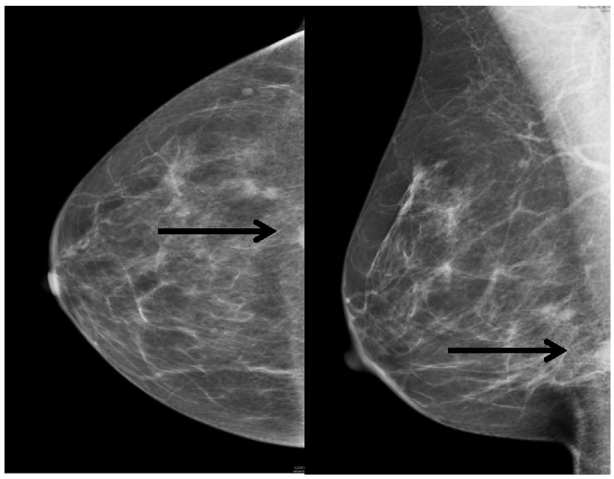

Figure 1. 15 mm indistinct mass partially seen at the very back of the right breast on both MLO and CC views. Core biopsy histology = Invasive ductal carcinoma Grade 3 ER/PR +ve HER 2−ve.

Phase 1

A questionnaire containing both closed and open questions aimed at assessing current practice in appraising mammographic IQ was initially sent to 2 UK breast screening units to test for content validity. No subsequent changes were made and it was then sent via on line Survey Monkey to centres in countries with a national mammography screening programme. In addition to establishing current practice, responses highlighted which centres had 4 radiographers meeting the inclusion criteria i.e. with a minimum of 4 years' experience in performing (not reporting) screening mammograms that could be taken forward for participation in phase 4 of the study.

Any information from the questionnaire requiring further elucidation was followed up by skype/telephone interview then all data transferred to MS Excel (Microsoft Corp, Redmond WA).

An expert panel was assembled from the respondents, inclusion criteria being a breast radiologist, or radiographer/other professional who either trains radiographers in undertaking mammography or has published work in peer reviewed journals investigating assessment of mammographic IQ and the associated criteria used.

Phase 2

The first panel task was to develop a list of criteria and their definitions to be used in the new assessment tool, then for each criterion, its level of importance in the assessment process. Members were individually sent a preliminary list of suggested criteria largely derived from UK PGMI guidelines. The panel was asked to consider several aspects of a revised system, first the *inclusion* and *wording* of the criteria using a Likert scale of 1–5 where 5 is complete agreement and 1 is no agreement. They were also invited to add any further criteria they felt should be included. A modified Delphi like process was used repeatedly to adjust responses until consensus was reached. The Delphi process was repeated to score the level of *importance* of each newly agreed criterion. In any cases of poor 'importance' agreement the criterion was either dropped from the list or sent back to the panel for re wording. The mean importance score for each criterion was calculated. Finally the panel was asked to consider categories and whether or not the PGMI categories should be replaced or altered.

Phase 3

To test agreement within the expert panel in interpreting the new criteria a mini test set of digital mammograms from 12 consecutively screened women aged 50–70 years of age was compiled by the principal investigator independent of the expert panel. Each case comprised four images (2x Cranial Caudal, 2x Medio Lateral Oblique views). The images were anonymised and numbered and then enriched with mammograms demonstrating a range of image quality flaws. Women were excluded if they had previously undergone breast cancer surgery, had implants, only one breast or a pacemaker.

To enable all participants to view and score the study images a web based image collection and annotation software developed as part of the OPTIMAM project was used.[14] The test set was uploaded onto the OPTIMAM server only accessible initially to the expert panel. On the advice of our institutions Research and Development department, we did not require Ethical or Trust approval to use NHS staff in this research. As all images used were anonymised and

Please cite this article in press as: 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 (2017), http://dx.doi.org/10.1016/j.radi.2017.03.004