

Assessment of Jordanian Radiologist Performance in the Detection of Breast Cancers

Mohammad Rawashdeh^{1*}, Mostafa Abdelrahman¹, Maha Zaitoun¹,
Mark F. McEntee², Kriscia Tapia², Patrick Brennan²

¹Faculty of Applied Medical Sciences, Jordan University of Science and Technology, Irbid, Jordan

²Medical Image Optimisation and Perception Group (MIOPeG), The Brain and Mind Centre, Faculty of Health Sciences, The University of Sydney, Sydney, Australia

Email: *marawashdeh@just.edu.jo

How to cite this paper: Rawashdeh, M., Abdelrahman, M., Zaitoun, M., McEntee, M.F., Tapia, K. and Brennan, P. (2018) Assessment of Jordanian Radiologist Performance in the Detection of Breast Cancers. *Open Journal of Medical Imaging*, 8, 41-53.

<https://doi.org/10.4236/ojmi.2018.83006>

Received: July 23, 2018

Accepted: September 3, 2018

Published: September 6, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study aims to monitor diagnostic accuracy amongst Jordanian mammography readers and identify parameters linked to higher levels of performance. In this study, we have used the Breast Screen Reader Assessment Strategy (BREAST) platform to facilitate 27 radiologists in reading a case set of 60 digital mammograms, 20 of which included cancers. Each case consisted of the four standard cranio-caudal (CC) and medio-lateral oblique (MLO) projections. All radiologists were registered to read mammograms at their workplace by the Jordanian Ministry of Health. Each reader was asked to locate any malignancies, provide a confidence rating using a scale of 1 - 5, and identify the type of appearance. All images were displayed using 8 MP monitor, supported by radiology workstations with full image manipulation facilities. Results were evaluated using Jackknife Alternative Free-Response Receiver Operating Characteristic (JAFROC). Demographics obtained from each radiologist regarding their experience, qualifications and breast-reading activities were correlated against JAFROC scores using Spearman techniques. The results showed that the mean JAFROC score was 0.52 (95% confidence interval (CI): 0.46, 0.58); location sensitivity score was 0.41 (95% CI: 0.41, 0.56); specificity score was 0.73 (95% CI: 0.68, 0.83). Higher performance in term of JAFROC scores was directly related to number of years since professional qualification ($r = 0.433$; $p = 0.024$), number of years reading breast images ($r = 0.62$; $r = 0.001$) and number of mammography images read per year ($r = 0.69$; $p = 0.001$). On the other hand, higher performance was inversely linked to the frequency of reading other modalities per week ($r = -0.48$; $p = 0.010$). No other statistical differences were significant. Finally, higher radiologists' performance in cancer detection is correlated with increasing the number of mammograms reads per week.

Keywords

Radiologist, Performance, Breast Cancers, JAFROC

1. Introduction

Wide inter-reader variations exist in cancer diagnosis with mammography [1] [2] [3] [4]. The retrospective identification of cancers missed [1] and inter-reader variations in cancer diagnosis reported both clinically [1] and in experimental studies [3] demonstrates that human factors are a major limitation to consistent outcomes for imaging modalities [5] [6] [7] [8]. Inter-reader variation and its associated errors can result in false negatives [9] [10], false positives [11] [12] [13] and over-diagnosis [14] [15]. False negative diagnosis prevents early detection and treatment of cancer, which may negatively impact upon survival outcomes [9]. False positive diagnosis has been shown to cause patient anxiety and results in additional examination and cost [12]. Over-diagnosis of the disease may result in overtreatment [16] [17], which may further expose patients to risk from ionizing radiation and treatment [17] [18]. Evidence is available that early diagnosis of breast cancer is associated with 30% - 40% reduction in mortality from the disease [17] [18]. Improving reader performance may also reduce recommendations for further diagnostic work-up such as additional imaging and biopsy and lower the cost of screening for breast cancer. Therefore, it is important to identify strategies to improve early detection and characterization of breast cancer with mammograms, and to improve reader performance in the diagnosis of the disease. The current work aims to identify factors that may improve reader performance and potentially improve the ability of radiologists to detect and characterize lesions on mammograms.

The literature demonstrates considerable degree of radiologists' errors and inter-radiologists' variability in mammography interpretation [19] [20] [21] [22]. Studies have shown that the proportion of breast cancer missed on mammography range from 1.3% to 39% [23] [24] [25]. Depending on the type and radiographic presentation of cancer, error rates may increase to 45%, and are common with subtle mammographic lesions such as architectural distortion [26] [27] [28]. Furthermore, some lesions may be visible in a mammogram and seen by radiologists, but may be overlooked because they are atypical. Thus, substantial proportions of missed or unreported malignant lesions can be seen on mammograms retrospectively [29] [30]. Even when malignant lesions are visible, some breast readers dismiss them due to insufficient prompts generated by such lesions or variability in knowledge and perceptions of readers with regards to the prompts [31] [32] [33]. Therefore, reader factors arise not only because of inadequate search, but also due to perceptual and decision-making errors [31] [32] [33]. Thus, the variability in search, perceptual, and decision-making patterns of radiologists may also be responsible for the wide inter-reader variability in de-

tection and characterization of potentially visible breast cancer as benign or malignant [31] [34] [35] [36] [37] [38].

Inter-reader variability in mammography interpretation has been shown to be a global phenomenon [39] [40], and underlines the need for practical approaches to improve cancer detection using mammography, including technological factors, reader characteristics, and other interventions. An understanding of parameters that limit breast cancer detection with mammography and ways of improving mammography performance may be crucial to reducing false positive and false negative diagnoses as well as inter-reader variability. This will in turn facilitate early treatment and further reduce mortality from the disease [41]. Whilst previous research [42]-[52] investigated the relationship between radiologists' performance and readers characteristics in UK and USA and Australia, this work will measure for the first time Jordanian reader performance in reading mammography and will determine whether the key readers characteristics that increase the detection of breast cancer are the same as previously reported. The data should contribute insights towards an improvement to the service women receive and help reduce radiology reporting variability in the future.

2. Methods

Institutional ethics review board approval was obtained (Grant No. 20170326). This study was conducted in Amman, Jordan.

2.1. Image Set

The test set comprised 60 mammograms cases, comprising a total number of 240 images, each case consisting of four images: left and right caudal cranial (CC) and mediolateral oblique (MLO) projections for each breast.

Twenty of the cases had biopsy-proven cancer, either ductal carcinoma in situ or invasive cancer with four of these cases containing multiple lesions. The forty remaining images were normal confirmed by follow up mammograms produced two years later. The normal cases contained incidental benign findings including calcified duct ectasia, calcified oil cysts, benign calcified fibro adenoma and intramammary lymph nodes.

2.2. Radiologist's Experience Details

A total group of 27 board-certified radiologists randomly participated in this study. Self-reported experience parameters including age, number of years since qualification as a radiologist, number of years reading mammograms, number of mammograms read per year and number of hours reading mammograms per week were recorded (Table 1).

2.3. Test Environment

Radiologists interpreted the images in a room 180 m² and with walls painted in light grey and brown matte colours to minimize specular reflection. A built-in

Table 1. Mean, Standard deviation (sd) for years certified, years reading mammograms, number of mammogram per year, number of mammogram per week, others modality score along with upper and lower 95% CI of mean.

| | Mean | sd | Lower 95% CI | Upper 95% CI |
|---|--------|--------|--------------|--------------|
| Number of years since professional qualification | 7.93 | 5.95 | 5.57 | 10.28 |
| Number of years reading breast images | 6.33 | 4.96 | 4.37 | 8.30 |
| Number of years reading breast images per year | 681.48 | 642.13 | 427.4 | 935.5 |
| Number of years reading breast images per week | 10.41 | 11.24 | 5.96 | 14.86 |
| Number of years reading breast images per session | 6.81 | 10.96 | 2.48 | 11.15 |

Integrated Front Sensor (IFS) measures brightness and gray scale tones to calibrate to DICOM Part 14. A calibrated photometer (Model Konica Minolta CL-200, Ramsey, NJ) was used to assess ambient light, which was maintained around 20 - 30 lux. Specifications of the workstation used for the work, such as monitor model, size, video card and calibration are described in **Table 2**.

2.4. Study Description

Radiologists were asked to localize and assess breast abnormalities according to the BI-RADS assessment categories used in Australia. The software platform used in the test was the Breast Reader Assessment Strategy (BREAST), which permits reading of digital images, determining of lesion location and providing an assessment category for breast lesions. The assessment categorization involved giving any perceived lesion a score of 2 (benign), 3 (equivocal), 4 (suspicious) and 5 (malignant). No information concerning the number of abnormal or normal cases was provided and the test software was explained to all radiologists before commencing the test. No time limit was imposed for the assessment of images and radiologists could freely access the panning, zooming and windowing post-processing tools. After a decision had been reached, radiologists located any perceived lesion, using a mouse-controlled cursor, on a laptop that simultaneously presented the same image as the one displayed on the high resolution monitors. If the decision about the case were “normal”, radiologists could just click on “next case” and the category score 1 (negative) would automatically appear for this case.

The web-based software provided general instructions on the process of reviewing, lesion marking and rating of the mammograms. Information on confidence level ratings to be used in the study was also provided to the readers. A short survey was included as part of the software to gather some general details on the participants’ demographic and clinical involvement. Overall demonstration of the software was given to each reader before the start of any readings. This platform allows radiologists to assess a mammographic test-set and obtain

Table 2. Shows workstation specifications.

| Parameters | Workstation |
|---------------------------|---|
| Type | EIZO, Japan |
| Monitor | RADI FORCE 850 |
| Monitor size | 8 megapixel |
| Display colors | 10-bit colors (display port): 1.07 billion (maximum) colors 8-bit colors: 16.77 million from a palette of 68 billion colors |
| Calibration | Digital imaging and communication in medicine |
| Standard display function | Gray-scale |
| Minimum luminance | 1.3 cd/m ² |
| Maximum luminance | 500 cd/m ² |
| Contrast ratio (typical) | 1450:1 |
| Number of workstations | 4 |

feedback on their performance, with the radiologists' correct decision, and errors made matched against the truth as shown in **Figure 1**.

2.5. Data Analysis

The numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each reader were counted. Sensitivity and specificity were then calculated. Sensitivity was calculated by dividing the number of TPs by the sum of TPs and FNs ($TP/(TP + FN)$). Specificity was calculated as a ratio of TN and the sum of FP and TN ($TN/(FP + TN)$). We also calculated location sensitivity (the proportion of true positives marked in the correct location as defined by a 75 pixel radius from the centre of the lesion). Jackknife Alternative Free-Response Receiver Operating Characteristic (JAFROC) software (Version 4.1) was used to calculate JAFROC figure of merit (FOM) values. A power analysis showed that with the sample size used in this study (60 cases and 27 radiologists) the detectable differences were 0.04, 0.07, and 0.05 for JAFROC, location sensitivity, and specificity, respectively, at 80% power.

Radiologists' performance was calculated using the previous metrics and correlated against key reader characteristics such as experience, qualifications, frequency of reading other modalities per week, and breast reading practices using Spearman techniques. Further analysis included a stepwise linear regression to predict the independent effect of the significant findings of the radiologist's experiences on JAFROC scores.

Additional analyses were performed to further assess key characteristics for specific mammographic reading volumes by categorizing readers in two sub-groups on the basis of the number of mammographic readings per year: fewer than 500, and more than 500. JAFROC data, location sensitivity, and specificity and compared using the t-test.

All statistical analyses were performed using the software IBM SPSS Statistics

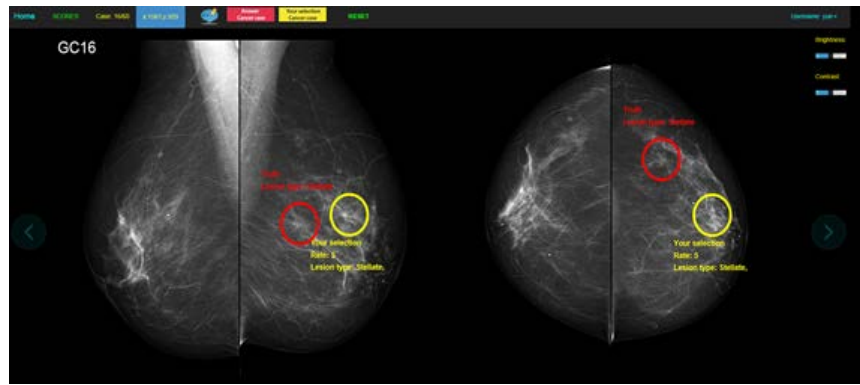


Figure 1. Example of BREAST interface showing readers' selection and the true location of cancer within the breast.

(version 22.0, for MAC; SPSS). Results were considered statistically significant when the p-value was ≤ 0.05 .

3. Results

Mean JAFROC, location sensitivity and specificity scores across all 27 readers are shown in **Table 3**, along with upper and lower 95% confidence intervals of the mean.

Higher performance in term of JAFROC scores was directly related to number of years since professional qualification ($r = 0.433$; $p = 0.024$), number of years reading breast images ($r = 0.62$; $p = 0.001$) and number of mammography images read per year ($r = 0.69$; $p = 0.001$). On the other hand, higher performance was inversely linked to the frequency of reading other modalities per week ($r = -0.48$; $p = 0.010$). No other statistical differences were significant (**Table 4**).

The stepwise regression revealed for JAFROC that a combination of the positive predictor which number of mammography images read per year ($r^2 = 0.416$, $p = 0.001$) and the negative predictor which is frequency of reading other modalities per week ($r^2 = -0.608$, $p = 0.008$), as a group, were more accurately predictive of JAFROC than was either variable alone. The line equation was $JAFROC = 0.780 + (Y \cdot 0.009) - (H \cdot 0.003)$ where Y is Mammograms read per year, H is frequency of reading other modalities per week.

Compared with the 14 (52%) readers who always maintained a total interpretive volume of at least 500 mammograms per year, the 13 (48%) readers who consistently had volume less than 500 mammograms per year experienced an 11% reduction in JAFROC (**Table 5**).

4. Discussion

The variability in radiologists' performance when reading mammograms is a concern across both screening and diagnostic mammography. Identifying causal factors for this variability is a first step towards optimising diagnostic efficacy. It is generally accepted that experience of the radiologist is a determinant of performance. Training, number of years since qualification, years of interpreting

Table 3. Mean TPs, specificity JAFROC, and location sensitivity specificity scores along with upper and lower 95% CI of mean.

| Score type | Mean | Lower 95% CI of mean | Upper 95% CI of mean |
|-----------------|------|----------------------|----------------------|
| % TPs | 0.71 | 0.64 | 0.78 |
| Specificity | 0.75 | 0.68 | 0.83 |
| JAFROC | 0.52 | 0.46 | 0.58 |
| % L sensitivity | 0.41 | 0.41 | 0.56 |

Table 4. Spearman correlation Analysis of the JAFROC, location sensitivity and specificity value with readers parameters are shown r values are shown in the table and p values are given in parentheses. Values shown in bold font are statistically significant.

| Score type | Values measures r (p) | | |
|--|-----------------------|----------------------|----------------|
| | JAFROC | Location sensitivity | Specificity |
| Age | -0.053 (0.792) | 0.179 (0.371) | -0.153 (0.447) |
| Years of qualification | 0.433 (0.024) | 0.147 (0.464) | 0.089 (0.659) |
| Years of mammogram reading | 0.628 (0.001) | 0.553 (0.003) | -0.045 (0.820) |
| Mammograms read per year | 0.696 (0.001) | 0.575 (0.002) | 0.046 (0.820) |
| Frequency of reading other modalities per week | -0.489 (0.01) | -0.273 (0.168) | -0.376 (0.053) |

Table 5. Correlation analysis of JAFROC, location sensitivity, and specificity values for radiologists with less and more readings than 500 per year (national requirement).

| | N | JAFROC | | Location sensitivity | | Specificity | |
|--------------------------|----|--------|---------|----------------------|---------|-------------|---------|
| | | Mean | p-value | Mean | p-value | Mean | p-value |
| Mammograms read per year | | | | | | | |
| <500 | 14 | 0.489 | 0.037 | 0.459 | 0.420 | 0.757 | 0.973 |
| >500 | 13 | 0.571 | | 0.523 | | 0.759 | |

mammograms and/or the number of mammograms read per year has been used as criteria for assessing radiologists' performance [42]. Many studies have assessed the impact of volume read per year in cancer detection with conflicting outcomes. Some studies have shown that volume read per year increases performance, and has potential for the optimization of screening mammography programs [42]-[48]. Other studies have reported decreased or no change in radiologists' performance irrespective of the volume read per year [49] [50] [51] [52].

The current work, investigated variations in diagnostic accuracy among readers who are currently involved in reporting breast images in Jordan. Higher levels of reader performance were found to be linked to numbers of years as certified radiologists, years of experience and hours readings per week.

The results of this work show that, although the number of cases read per year increased the ability of radiologists to correctly detect cancer in mammograms, it

did not prevent them from making false positive errors (reporting the presence of cancer where there is none). It has been shown that perception of cancer and diagnostic decision-making relies on the reader's previous reader knowledge and experience [53] [54]. Therefore, improvement in sensitivity could be attributed to increased exposure to a wide range of mammographic features of cancer from increased number of cases read. The heterogeneity of the breast parenchyma and the mimicking of cancer by normal tissue may be partly implicated in the higher number of false positives. Because of the medico legal implication of false-negative diagnosis [55], mean radiologists tend to report perturbations in the mammogram that may be suspicious of cancer and increase their recall rates. Additionally, radiologists who participated in the study reported in this thesis were assessed in a "laboratory" setting, not in their normal clinical setting. In such a setting, radiologists tend to expect more abnormal cases of cancer, prompting interpretation of normal parenchymal perturbations that are suspicious as cancer [56].

Although we found that increases in the number of mammography images read per year are associated with higher performance, previous work has shown widely varying results [3]-[13]. Such discrepancies in findings may be explained, at least in part, by different methods employed. In addition, most studies are based on a selected sample of radiologists [3] [5] [8] [11] [12] or excluded some radiologists on the basis of their experience or their volume [6] [7] [13]. Finally, some studies did not adjust for potential confounders [4] [9] such as ambient light and viewing conditions. The US and Canada have similar interpretive volume requirements of at least 480 mammograms per year [14] [15]. Our results provide evidence in support of this annual requirement.

Number of mammography images read per year is also associated with improvement in location sensitivity. Understanding whether mammography-screening accuracy can be affected by the degree of radiologist involvement of a radiologist in diagnostic investigation of abnormal screening mammograms, including imaging and biopsies, is an important question in need of further study.

One educational project, which offers readers educational experiences and feedback, is BREAST [57]. The matching of errors against the truth provides radiologists with feedback about the nature of lesions missed. They can review their correct and incorrect cases and thus learn from the feedback. This may in turn facilitate tailoring of training regimens to improve mammographic interpretation performance. Feedback on performance is also very useful to employers as well, enabling them to identify areas of need for further education of their employees. Previous research has demonstrated that test-set reading interventions like BREAST are useful as they provide immediate feedback on correct and incorrect diagnosis. It is hoped that through multiple interventions like BREAST the accuracy of mammography interpretation would significantly improve [58].

It should be acknowledged that there are limitations in the work. Firstly, the number of cases assessed was relatively small, and the case mix was not typical of

a screening environment having many more abnormal than would normally be expected. Also, prior cases were not included, which could have had some influence on the results.

In summary, radiologists' performance improves with increasing number of mammograms read per week, and by focusing their duties towards mammogram reading. The use of interventional educational programs, such as BREAST, could be applied to compensate for low reading volumes and help to expand radiological skills necessary to accurately identify breast patterns and lesions. The results have potential implications for breast screening efficacy and women's anxiety.

Acknowledgements

1) Appreciation and thanks are due to Jordan University of Sciences and Technology for their research grant (Grant No. 20170326). 2) The authors acknowledge the kind support of the Breast Screen Reader Assessment Strategy (BREAST) for providing a platform. 3) The authors would like to thank Jordan Breast Cancer Program for their support towards the completion of this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] Ciccone, G., Vineis, P., Frigerio, A., *et al.* (1992) Inter-Observer and Intra-Observer Variability of Mammogram Interpretation: A Field Study. *European Journal of Cancer*, **28**, 1054-1058. [https://doi.org/10.1016/0959-8049\(92\)90455-B](https://doi.org/10.1016/0959-8049(92)90455-B)
- [2] Drukker, K., Horsch, K.J., Pesce, L.L., *et al.* (2013) Interreader Scoring Variability in an Observer Study Using Dual-Modality Imaging for Breast Cancer Detection in Women with Dense Breasts. *Academic Radiology*, **20**, 847-853. <https://doi.org/10.1016/j.acra.2013.02.007>
- [3] Duijm, L.E., Louwman, M.W., Groenewoud, J.H., *et al.* (2009) Inter-Observer Variability in Mammography Screening and Effect of Type and Number of Readers on Screening Outcome. *British Journal of Cancer*, **100**, 901-907. <https://doi.org/10.1038/sj.bjc.6604954>
- [4] Jackson, S.L., Sickles, E.A., Abraham, L., *et al.* (2009) Variability of Interpretive Accuracy among Diagnostic Mammography Facilities. *Journal of the National Cancer Institute*, **101**, 814-827. <https://doi.org/10.1093/jnci/djp105>
- [5] Elmore, J.G., Jackson, S.L., Abraham, L., *et al.* (2009) Variability in Interpretive Performance at Screening Mammography and Radiologists' Characteristics Associated with Accuracy. *Radiology*, **253**, 641-651. <https://doi.org/10.1148/radiol.2533082308>
- [6] Elmore, J.G., Wells, C.K., Lee, C.H., *et al.* (1994) Variability in Radiologists' Interpretations of Mammograms. *The New England Journal of Medicine*, **331**, 1493-1499. <https://doi.org/10.1056/NEJM199412013312206>
- [7] Klompenhouwer, E.G., Duijm, L.E., Voogd, A.C., *et al.* (2014) Variations in Screening Outcome among Pairs of Screening Radiologists at Non-Blinded Double Reading of Screening Mammograms: A Population-Based Study. *European Radiol-*

- ogy, **24**, 1097-1104. <https://doi.org/10.1007/s00330-014-3102-4>
- [8] Pan, H.B., Yang, T.L., Hsu, G.C., et al. (2012) Can Missed Breast Cancer Be Recognized by Regular Peer Auditing on Screening Mammography? *Journal of the Chinese Medical Association*, **75**, 464-467. <https://doi.org/10.1016/j.jcma.2012.06.018>
- [9] Burrell, H.C., Evans, A.J., Wilson, A.R., et al. (2001) False-Negative Breast Screening Assessment: What Lessons Can We Learn? *Clinical Radiology*, **56**, 385-388. <https://doi.org/10.1053/crad.2001.0662>
- [10] Huynh, P.T., Jarolimek A.M. and Daye S. (1998) The False-Negative Mammogram. *RadioGraphics*, **18**, 1137-1154. <https://doi.org/10.1148/radiographics.18.5.9747612>
- [11] Bond, M., Pavey, T., Welch, K., et al. (2013) Psychological Consequences of False-Positive Screening Mammograms in the UK. *Evidence-Based Medicine*, **18**, 54-61. <https://doi.org/10.1136/eb-2012-100608>
- [12] Consedine, N.S. (2013) A False-Positive on Screening Mammography Has a Negative Psychosocial Impact up to 3 Years after Receiving the All Clear. *Evidence-Based Mental Health*, **16**, 115. <https://doi.org/10.1136/eb-2013-101410>
- [13] Espasa, R., Murta-Nascimento, C., Bayes, R., et al. (2012) The Psychological Impact of a False-Positive Screening Mammogram in Barcelona. *Journal of Cancer Education*, **27**, 780-785. <https://doi.org/10.1007/s13187-012-0349-9>
- [14] Coldman, A. and Phillips, N. (2013) Incidence of Breast Cancer and Estimates of Overdiagnosis after the Initiation of a Population-Based Mammography Screening Program. *Canadian Medical Association Journal*, **185**, E492-E498. <https://doi.org/10.1503/cmaj.121791>
- [15] Duffy, SW., Agbaje, O., Tabar, L., et al. (2005) Overdiagnosis and Overtreatment of Breast Cancer: Estimates of Overdiagnosis from Two Trials of Mammographic Screening for Breast Cancer. *Breast Cancer Research*, **7**, 258-265. <https://doi.org/10.1186/bcr1354>
- [16] de Koning, H.J., Draisma, G., Fracheboud, J., et al. (2006) Overdiagnosis and Overtreatment of Breast Cancer: Microsimulation Modelling Estimates Based on Observed Screen and Clinical Data. *Breast Cancer Research*, **8**, 202. <https://doi.org/10.1186/bcr1369>
- [17] Gunsoy, N.B., Garcia-Closas, M. and Moss, S.M. (2014) Estimating Breast Cancer Mortality Reduction and Overdiagnosis Due to Screening for Different Strategies in the United Kingdom. *British Journal of Cancer*, **110**, 2412-2419. <https://doi.org/10.1038/bjc.2014.206>
- [18] Njor, S., Nystrom, L., Moss, S., et al. (2012) Breast Cancer Mortality in Mammographic Screening in Europe: A Review of Incidence-Based Mortality Studies. *Journal of Medical Screening*, **19**, 33-41. <https://doi.org/10.1258/jms.2012.012080>
- [19] Miglioretti, D.L., Ichikawa, L., Smith, R.A., Bassett, L.W., Feig, S.A., Monsees, B., Carney, P.A., et al. (2015) Criteria for Identifying Radiologists with Acceptable Screening Mammography Interpretive Performance Based on Multiple Performance Measures. *American Journal of Roentgenology*, **204**, W486-W491. <https://doi.org/10.2214/AJR.13.12313>
- [20] Bezic, J., Mrklic, I., Pogorelic, Z., et al. (2013) Mammographic Screening Has Failed to Improve Pathohistological Characteristics of Breast Cancers in Split Region of Croatia. *Breast Disease*, **34**, 47-51. <https://doi.org/10.3233/BD-130349>
- [21] Kelly, K.M., Dean, J., Lee, S.J., et al. (2010) Breast Cancer Detection: Radiologists' Performance Using Mammography with and without Automated Whole-Breast Ultrasound. *European Radiology*, **20**, 2557-2564.

- <https://doi.org/10.1007/s00330-010-1844-1>
- [22] Pitman, A.G., Tan, S.Y., Ong, A.H., et al. (2011) Intrareader Variability in Mammographic Diagnostic and Perceptual Performance amongst Experienced Radiologists in Australia. *Journal of Medical Imaging and Radiation Oncology*, **55**, 245-251. <https://doi.org/10.1111/j.1754-9485.2011.02260.x>
- [23] Evans, K.K., Birdwell, R.L. and Wolfe, J.M. (2013) If You Don't Find It often, You often Don't Find It: Why Some Cancers Are Missed in Breast Cancer Screening. *PLoS ONE*, **8**, e64366. <https://doi.org/10.1371/journal.pone.0064366>
- [24] Hofvind, S., Skaane, P., Vitak, B., et al. (2005) Influence of Review Design on Percentages of Missed Interval Breast Cancers: Retrospective Study of Interval Cancers in a Population-Based Screening Program. *Radiology*, **237**, 437-443. <https://doi.org/10.1148/radiol.2372041174>
- [25] Maxwell, A.J. (1999) Breast Cancers Missed in the Prevalent Screening Round: effect upon the Size Distribution of Incident round Detected Cancers. *Journal of Medical Screening*, **6**, 28-29. <https://doi.org/10.1136/jms.6.1.28>
- [26] Banik, S., Rangayyan, R.M. and Desautels, J.E. (2009) Detection of Architectural Distortion in Prior Mammograms of Interval-Cancer Cases with Neural Networks. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference*, 6667-6670. <https://doi.org/10.1109/IEMBS.2009.5334517>
- [27] Banik, S., Rangayyan, R.M. and Desautels, J.E. (2011) Detection of Architectural Distortion in Prior Mammograms. *IEEE Transactions on Medical Imaging*, **30**, 279-294. <https://doi.org/10.1109/TMI.2010.2076828>
- [28] Bird, R.E., Wallace, T.W. and Yankaskas, B.C. (1992) Analysis of Cancers Missed at Screening Mammography. *Radiology*, **184**, 613-617. <https://doi.org/10.1148/radiology.184.3.1509041>
- [29] Mello-Thoms, C., Dunn, S., Nodine, C.F., et al. (2002) The Perception of Breast Cancer: What Differentiates Missed from Reported Cancers in Mammography? *Academic Radiology*, **9**, 1004-1012. [https://doi.org/10.1016/S1076-6332\(03\)80475-0](https://doi.org/10.1016/S1076-6332(03)80475-0)
- [30] Mello-Thoms, C., Dunn, S.M., Nodine, C.F., et al. (2001) An Analysis of Perceptual Errors in Reading Mammograms Using Quasi-Local Spatial Frequency Spectra. *Journal of Digital Imaging*, **14**, 117-123. <https://doi.org/10.1007/s10278-001-0010-3>
- [31] Mello-Thoms, C. (2006) The Problem of Image Interpretation in Mammography: Effects of Lesion Conspicuity on the Visual Search Strategy of Radiologists. *British Journal of Radiology*, **79**, S111-S116. <https://doi.org/10.1259/bjr/61144371>
- [32] Kundel, H.L. (1982) Disease Prevalence and Radiological Decision Making. *Investigative Radiology*, **17**, 107-109. <https://doi.org/10.1097/00004424-198201000-00020>
- [33] Kundel, H.L. and Nodine, C.F. (1983) A Visual Concept Shapes Image Perception. *Radiology*, **146**, 363-368. <https://doi.org/10.1148/radiology.146.2.6849084>
- [34] Mello-Thoms, C. (2006) How Does the Perception of a Lesion Influence Visual Search Strategy in Mammogram Reading? *Academic Radiology*, **13**, 275-288. <https://doi.org/10.1016/j.acra.2005.11.034>
- [35] Nodine, C.F., Kundel, H.L., Lauer, S.C., et al. (1996) Nature of Expertise in Searching Mammograms for Breast Masses. *Academic Radiology*, **3**, 1000-1006. [https://doi.org/10.1016/S1076-6332\(96\)80032-8](https://doi.org/10.1016/S1076-6332(96)80032-8)
- [36] Nodine, C.F., Kundel, H.L., Mello-Thoms, C., et al. (1999) How Experience and Training Influence Mammography Expertise. *Academic Radiology*, **6**, 575-585. [https://doi.org/10.1016/S1076-6332\(99\)80252-9](https://doi.org/10.1016/S1076-6332(99)80252-9)

- [37] Torres-Mejia, G., Villasenor-Navarro, Y., Yunes-Diaz, E., et al. (2011) Validity and Reliability of Mammographic Interpretation by Mexican Radiologists, Using the BI-RADS System. *Revista de Investigacion Clinica; organo del Hospital de Enfermedades de la Nutricion*, **63**, 124-134.
- [38] Abdullah, N., Mesurole, B., El-Khoury, M., et al. (2009) Breast Imaging Reporting and Data System Lexicon for US: Interobserver Agreement for Assessment of Breast Masses. *Radiology*, **252**, 665-672. <https://doi.org/10.1148/radiol.2523080670>
- [39] Elverici, E., Zengin, B., Nurdan, B.A., et al. (2013) Interobserver and Intraobserver Agreement of Sonographic BIRADS Lexicon in the Assessment of Breast Masses. *Iranian Journal of Radiology*, **10**, 122-127. <https://doi.org/10.5812/iranradiol.10708>
- [40] Lazarus, E., Mainiero, M.B., Schepps, B., et al. (2006) BI-RADS Lexicon for US and Mammography: Interobserver Variability and Positive Predictive Value. *Radiology*, **239**, 385-391. <https://doi.org/10.1148/radiol.2392042127>
- [41] Berg, W.A. (2010) Benefits of Screening Mammography. *JAMA*, **303**, 168-169. <https://doi.org/10.1001/jama.2009.1993>
- [42] Rawashdeh, M.A., Lee, W.B., Bourne, R.M., et al. (2013) Markers of Good Performance in Mammography Depend on Number of Annual Readings. *Radiology*, **269**, 61-67. <https://doi.org/10.1148/radiol.13122581>
- [43] Esserman, L., Cowley, H., Eberle, C., et al. (2002) Improving the Accuracy of Mammography: Volume and Outcome Relationships. *Journal of the National Cancer Institute*, **94**, 369-375. <https://doi.org/10.1093/jnci/94.5.369>
- [44] Jensen, A., Vejborg, I., Severinsen, N., et al. (2006) Performance of Clinical Mammography: A Nationwide Study from Denmark. *International Journal of Cancer*, **119**, 183-191. <https://doi.org/10.1002/ijc.21811>
- [45] Reed, W.M., Lee, W.B., Cawson, J.N., et al. (2010) Malignancy Detection in Digital Mammograms: Important Reader Characteristics and Required Case Numbers. *Academic Radiology*, **17**, 1409-1413. <https://doi.org/10.1016/j.acra.2010.06.016>
- [46] Kan, L., Olivotto, I.A., Warren, Burhenne, L.J., et al. (2000) Standardized Abnormal Interpretation and Cancer Detection Ratios to Assess Reading Volume and Reader Performance in a Breast Screening Program. *Radiology*, **215**, 563-567. <https://doi.org/10.1148/radiology.215.2.r00ma42563>
- [47] Elmore, J.G., Wells, C.K. and Howard, D.H. (1998) Does Diagnostic Accuracy in Mammography Depend on Radiologists' Experience? *Journal of Women's Health*, **7**, 443-449.
- [48] Smith-Bindman, R., Chu, P., Miglioretti, D.L., et al. (2005) Physician Predictors of Mammographic Accuracy. *Journal of the National Cancer Institute*, **97**, 358-367. <https://doi.org/10.1093/jnci/dji060>
- [49] Barlow, W.E., Chi, C., Carney, P.A., et al. (2004) Accuracy of Screening Mammography Interpretation by Characteristics of Radiologists. *Journal of the National Cancer Institute*, **96**, 1840-1850. <https://doi.org/10.1093/jnci/djh333>
- [50] Beam, C.A., Conant, E.F. and Sickles, E.A. (2003) Association of Volume and Volume-Independent Factors with Accuracy in Screening Mammogram Interpretation. *Journal of the National Cancer Institute*, **95**, 282-290. <https://doi.org/10.1093/jnci/95.4.282>
- [51] Cornford, E., Reed, J., Murphy, A., et al. (2011) Optimal Screening Mammography Reading Volumes; Evidence from Real Life in the East Midlands Region of the NHS Breast Screening Programme. *Clinical Radiology*, **66**, 103-107. <https://doi.org/10.1016/j.crad.2010.09.014>

- [52] Haneuse, S., Buist, D.S., Miglioretti, D.L., *et al.* (2012) Mammographic Interpretive Volume and Diagnostic Mammogram Interpretation Performance in Community Practice. *Radiology*, **262**, 69-79. <https://doi.org/10.1148/radiol.11111026>
- [53] Al-Khawari, H., Athyal, R.P., Al-Saeed, O., *et al.* (2010) Inter- and Intraobserver Variation between Radiologists in the Detection of Abnormal Parenchymal Lung Changes on High-Resolution Computed Tomography. *Annals of Saudi Medicine*, **30**, 129-133. <https://doi.org/10.4103/0256-4947.60518>
- [54] Elmore, J.G., Taplin, S.H., Barlow, W.E., *et al.* (2005) Does Litigation Influence Medical Practice? The Influence of Community Radiologists' Medical Malpractice Perceptions and Experience on Screening Mammography. *Radiology*, **236**, 37-46. <https://doi.org/10.1148/radiol.2361040512>
- [55] Nodine, C.F., Mello-Thoms, C., Kundel, H.L., *et al.* (2002) Time Course of Perception and Decision Making during Mammographic Interpretation. *AJR*, **179**, 917-923. <https://doi.org/10.2214/ajr.179.4.1790917>
- [56] Gale, A.G. (2003) Performs: A Self-Assessment Scheme for Radiologists in Breast Screening. *Seminars in Breast Disease*, **6**, 148-152. <https://doi.org/10.1053/j.sembd.2004.03.006>
- [57] Brennan, P.C., Ryan, J. and Lee, W. (2013) BREAST: A Novel Method to Improve the Diagnostic Efficacy of Mammography In: Mello-Thoms, C.K.A.C.R., Ed., *SPIE Medical Imaging*, Medical Imaging, Florida, 867307.
- [58] Suleiman, W.I., Rawashdeh, M.A., Lewis, S.J., McEntee, M.F., Lee, W. and Tapia, K. (2016) Impact of Breast Reader Assessment Strategy on Mammographic Radiologists' Test Reading Performance. *Journal of Medical Imaging and Radiation Oncology*, **60**, 352-358. <https://doi.org/10.1111/1754-9485.12461>