Markers of Good Performance in Mammography Depend on Number of Annual Readings[1]

Mohammad A. Rawashdeh, MSc
Warwick B. Lee, FRANZCR
Roger M. Bourne, PhD
Elaine A. Ryan, PhD
Mariusz W. Pietrzyk, PhD
Warren M. Reed, PhD
Robert C. Heard, PhD
Deborah A. Black, PhD
Patrick C. Brennan, PhD

Purpose: To explore relationships between reader performance and reader characteristics in mammography for specific radiologist groupings on the basis of annual number of readings.

Materials and Methods: The institutional review board approved the study and waived the need for patient consent to use all images. Readers gave informed consent. One hundred sixteen radiologists independently reviewed 60 mammographic cases: 20 cases with cancer and 40 cases with normal findings. Readers located any visualized cancer, and levels of confidence were scored from 1 to 5. A jackknifing free response operating characteristic (JAFROC) method was used, and figures of merit along with sensitivity and specificity were correlated with reader characteristics by using Spearman techniques and standard multiple regressions.

Results: Reader performance was positively correlated with number of years since qualification as a radiologist ($P \leq .01$), number of years reading mammograms ($P \leq .03$), and number of readings per year ($P \leq .0001$). The number of years since qualification as a radiologist ($P \leq .004$) and number of years of reading mammograms ($P \leq .002$) were negatively related to JAFROC values for radiologists with annual volumes of less than 1000 mammographic readings. For individuals with more than 5000 mammographic readings per year, JAFROC values were positively related to the number of years that the reader was qualified as a radiologist ($P \leq .01$), number of years of reading mammograms ($P \leq .002$), and number of hours per week of reading mammograms ($P \leq .003$). Number of mammographic readings per year was positively related with JAFROC scores for readers with an annual volume between 1000 and 5000 readings ($P \leq .03$). Differences in JAFROC scores appear to be more related to specificity than location sensitivity, with the former demonstrating significant relationships with four of the five characteristics analyzed, whereas no relationships were shown for the latter.

Conclusion: Radiologists' determinants of performance are associated with annual reading volumes. Ability to recognize normal images is a discriminating factor in individuals with a high volume of mammographic readings.

© RSNA, 2013

[1] From the Medical Image Optimisation and Perception Group (MIOPeG), Medical Imaging and Radiation Sciences, Faculty Research Group, Faculty of Health Sciences, the University of Sydney, Cumberland Campus, 75 East St, Lidcombe, NSW 2141, Australia (M.A.R., R.M.B., E.A.R., M.W.P., W.M.R., R.C.H., D.A.B., P.C.B.); and BreastScreen, Cancer Institute NSW, Alexandria, New South Wales, Australia (W.B.L.). Received November 20, 2012; revision requested January 4, 2013; final revision received February 25; accepted March 15; final version accepted April 1. Supported by the National Breast Cancer Foundation, the Australian Department of Health and Aging, and the Royal Australian and New Zealand College of Radiologists. Jordan University of Science and Technology provided the student scholarship. **Address correspondence to** M.A.R. (e-mail: *mraw2971@uni.sydney.edu.au*).

Technological advancements in mammography include the transition from screen-film to soft-copy reading and the development of computer-aided diagnostic systems (1); however, cancer detection still depends on an individual reader's interpretation (2–6). As long as radiologists interpret medical images, the human factor and its limitations on abnormality detection will continue to influence overall performance. Errors in reporting can be caused by multiple factors, but it is estimated that perceptual factors account for 60% of all diagnostic errors in radiology (7). A perceptual error is not a single entity but rather arises from three individual factors: inadequate knowledge to be able to search an image properly, inability to recognize an abnormality, and faulty decision-making as to whether an image entity is a normal variant or a pathologic lesion (8). Minimization of perceptual errors depends on the radiologist's characteristics, such as specialization and experience.

Previous researchers who focused on board-certified radiologists have looked at reader characteristics in terms of training and clinical experience; however, a number of unanswered questions remain (5,9–11). First, evidence is conflicting as to which are the key characteristics. While one author has highlighted fellowship training in breast imaging as a key parameter (12), others have stressed that higher performance is linked to levels of experience as represented by years since certification, years of experience of reading mammograms, and hours of reading per week (9,10). Second, the need for radiologists to read an annual minimum number of mammographic images has not been consistently justified, even though a number of countries now state minimal levels (11,13–15). Barlow et al (5) and Miglioretti et al (9) demonstrated no relationship between the number of reads and performance, in contrast to the strong correlations reported by Reed et al (10). These differing results highlight the need for a more in-depth analysis on annual readings (10).

The receiver operating characteristic method has been the main method for most of the observer performance studies reported so far (16). In studies with the receiver operating characteristic, the reader views and classifies each image as normal or abnormal (17), an approach that does not consider lesion location or allow for the identification of multiple abnormalities per image. The work reported here uses the jackknifing free response operating characteristic (JAFROC) method, which addresses lesion location and multiplicity and is more representative of how radiologists actually interpret images in clinical practice (18).

Through use of a large number of breast image readers, the aim of the current study was to explore relationships between reader performance and reader characteristics in mammography for specific radiologist groupings on the basis of the annual number of readings.

Advances in Knowledge

- Factors linked to poor performance in mammographic reporting depend on volume of mammographic readings per year.
- The performance of individuals who read fewer than 1000 mammograms per year decreased with an increasing number of years of reading mammograms ($P \leq .002$).
- The variance among those who read more than 5000 mammograms per year is linked to the ability to recognize normal images.

Implication for Patient Care

- Identifying causal agents for variability in the performance of reading mammograms is a first step toward optimizing diagnostic effectiveness.

Materials and Methods

The institutional review board approved the study and waived the need for patient consent for the use of all images. Readers gave informed consent. The work was performed at the 2011 Royal Australian and New Zealand College of Radiologists Breast Imaging meeting in Hobart, Tasmania, Australia, between February 28 and March 4, 2011.

Observer Population

A total of 116 board-certified radiologists participated in the study, and self-reported demographic details in regard to their age, the number of years of experience after certification, the number of screening mammographic readings per year, and the number of hours per week of reading screening mammographic images are shown in Table 1. The mean age of readers was 50.7 years, the mean number of years since qualification as a radiologist was 14.4, and the mean number of years of experience in reading mammograms was 10.9 years . The mean number of hours per week spent reading breast images was 12.2, with a mean of 4005 mammographic readings per year.

Image Test Set

A test set containing 60 digital mammograms was presented in the same order to each observer. Each case in a patient consisted of four images: a caudal cranial and a mediolateral oblique projection of each breast. The cases were chosen by one of the authors

Published online before print
10.1148/radiol.13122581 **Content codes:** BR DM

Radiology 2013; 269:61–67

Abbreviation:
JAFROC = jackknifing free response operating characteristic

Author contributions:
Guarantor of integrity of entire study, M.A.R.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, M.A.R., W.M.R., P.C.B.; clinical studies, M.A.R., W.B.L., R.M.B., E.A.R., W.M.R., P.C.B.; statistical analysis, M.A.R., E.A.R., M.W.P., W.M.R., R.C.H., D.A.B., P.C.B.; and manuscript editing, all authors

Conflicts of interest are listed at the end of this article.

Table 1

Details on 116 Participating Readers, with Interquartile Ranges

Parameter	Mean	First Quartile	Third Quartile	Standard Deviation	Interquartile Range
Age (y)	50.7	44	57	9	31–75
No. of years since qualification as a radiologist	14.4	8	20	8.4	1–42
No. of years of reading mammograms	10.9	4	17	7	1–30
No. of mammographic readings per year	4005	1500	5000	4076	200–25 000
No. of hours per week of reading mammograms	12.2	3	20	11.3	1–40

(W.B.L., a senior radiologist responsible for assessing training, support, and quality and clinical policies of BreastScreen New South Wales) on the basis that they were typical screening cases. The test set was enriched with cases with abnormal findings. There were 40 cases with normal findings and 20 cases with abnormal findings; the former were confirmed by two independent normal screen readings, review of the cases by two expert radiologists, and a follow-up negative screening mammogram obtained 2 years later. The cases with normal findings were a mixture of cases with "pristine" normal findings and cases with incidental benign lesions, such as calcified oil cysts, intramammary lymph nodes, calcified duct ectasia, and typically benign calcified fibroadenomas. The cases with "pristine" normal findings were chosen to reflect a normal screening caseload with regard to mammographic density. All 20 abnormal images contained a biopsy-proved malignancy (invasive cancer or ductal carcinoma in situ), with single and multicentric cancer in 16 and four images, respectively; 16 images contained mass lesions only, and four demonstrated calcifications. The level of difficulty ranged from subtle to difficult, with the percentage of readers detecting individual lesions varying from 33% to 95%.

Reading Environments

To match the radiologic reporting environment, readings were conducted in two reading rooms that measured 60 m^2 and 90 m^2, respectively, with ambient lighting of 12–20 lux at the position of the reader, measured with a calibrated photometer (Model Konica Minolta CL-200; Konica Minolta, Ramsey, NJ). Walls were painted a light gray and brown matte color, with minimum specular reflection. All images were reviewed by using a workstation (Sectra, Linköping, Sweden; Hologic, Bedford, Mass), each driving a 5-megapixel reporting monitor; one room featured the Sectra workstation, and the other, Hologic. The first workstation was linked to a monitor (MFGD 5621; Barco, Kortrijk, Belgium) driven by a video card (BarcoMed 5MP2FH; Barco), and the second displayed images by using a monitor (RadiForce G51; Eizo, Ishikawa, Japan) with a different video card (Matrox MED5MP-DVI; Dorval, Quebec, Canada). All monitors were calibrated to the Digital Imaging and Communication in Medicine gray-scale standard display function and displayed a maximum luminance within 5% of 475 $cd \cdot m^{-2}$, minimum luminance of 1.3 $cd \cdot m^{-2}$, and a contrast ratio of 365:1. To confirm consistency between workstations, radiologist JAFROC scores were compared; no differences were seen. Readers had unlimited time and were able to move through the image bank as they desired and change case reports to correct any perceived errors they had made. Readers also had access to a normal range of postprocessing tools, including windowing, panning, and zooming.

Study Design

Readers were asked to assess images from the same 60 cases in patients to detect breast lesions on the images and to mark anywhere within the lesion with a confidence level ranging from one to five, with one representing complete confidence that the image was normal and five representing complete confidence that a malignant lesion was present. If the observer could not detect any lesions on an image, he or she marked it as normal (confidence rating 1) and proceeded to the next image. The software was demonstrated to all readers before any readings, and no information was provided on the number of cases with abnormal findings or the number of lesions on the images. Once readers had made a decision on the basis of each image displayed as above, and because it was not possible to load the scoring software (Ziltron Technologies, San Diego, Calif) onto the workstations, they recorded their interpretation and located any lesions on a laptop that displayed the same images as those displayed on the reporting monitors, with all cancers being visible.

Data Analysis

JAFROC analysis was performed, and specificity and location sensitivity were calculated for the whole group of radiologists. Specificity was defined by the fraction of cases with normal findings that were correctly identified. Location sensitivity was determined by the distance of the mouse click from the center of the lesion. If the distance was within 2 cm, a correct decision was given. The mean scores of the three reader groups on these measures were compared by using a one-way analysis of variance and with Scheffé follow-up tests. Spearman correlations were used to correlate resulting values with age, number of years since qualification as a radiologist, number of years of reading mammograms, number of mammographic readings per year, and number of hours per week of reading mammograms.

Additional analyses were performed to further assess key characteristics for specific mammographic reading

Table 2

Mean JAFROC, Location Sensitivity, and Specificity Values

Score Type	Mean
JAFROC	0.78 (0.77, 0.80)
Location sensitivity	0.56 (0.53, 0.59)
Specificity	0.72 (0.69, 0.75)

Note.—Numbers in parentheses are 95% confidence intervals of the mean.

Table 3

Correlation Analysis of JAFROC, Location Sensitivity, and Specificity Values with Reader Parameters by Using Nonparametric Spearman Methods

	JAFROC		Location Sensitivity		Specificity	
Parameter	*r* Value	*P* Value	*r* Value	*P* Value	*r* Value	*P* Value
Age	0.05	≤.58	−0.13	≤.13	0.11	≤.22
No. of years since qualification as a radiologist	0.22	≤.01*	0.034	≤.71	0.18	≤.04*
No. of years of reading mammograms	0.19	≤.03*	−0.01	≤.90	0.19	≤.04*
No. of mammographic readings per year	0.28	≤.0001*	0.18	≤.04*	0.18	≤.0001*
No. of hours per week of reading mammograms	0.08	≤.38	0.09	≤.31	0.07	≤.43

* Findings with a significant difference.

Table 4

Mean JAFROC, Location Sensitivity, and Specificity Values

	Mean		
Annual Volume of Readings	JAFROC	Location Sensitivity	Specificity
<1000	0.67 (0.62, 0.73)*	0.53 (0.45, 0.60)	0.60 (0.51, 0.70)†
1000–5000	0.79 (0.76, 0.81)*	0.57 (0.53, 0.61)	0.72 (0.68, 0.76)†
>5000	0.86 (0.83, 0.88)*	0.59 (0.54, 0.64)	0.83 (0.80, 0.87)*

Note.—Annual volume of readings refers to annual number of mammographic readings. Numbers in parentheses are the 95% confidence intervals of the mean.

* Values are significant from each other ($P < .001$), except where otherwise indicated by the dagger symbol.

† $P = .014$.

volumes by categorizing readers in three subgroups on the basis of the number of mammographic readings per year: fewer than 1000, 1000–5000, and more than 5000. The means and 95% confidence intervals of the mean were calculated for JAFROC data, location sensitivity, and specificity. With use of the Spearman test, the performance of radiologists in each of these subgroups was then correlated against the same radiologists' characteristics as described for the total group.

Significant univariate predictors ($P \leq .05$) were also entered for all three subgroups (<1000, 1000–5000, and >5000 mammographic readings per year) into standard multiple regressions for the JAFROC data after checking for multivariate outliers, normality, linearity, and homoscedasticity for each group. Regression R^2 values were tested for significance by using *F* ratios. For the readers with an annual volume of fewer than 1000 mammographic readings per year, one multivariate outlier who had a significant Mahalanobis distance was excluded. Software (SPSS, version 21.0; SPSS, Chicago, Ill) was used for all statistical computations. Results with a *P* value of .05 or less were deemed to represent significant differences.

Results

The JAFROC scores appeared to be driven largely by increases in specificity because the groups did not differ in sensitivity (Table 2). Correlation analysis demonstrated that, for JAFROC and specificity values, three key characteristics were significantly linked to the radiologist's performance: number of years since qualification as a radiologist ($P \leq .01$ and $P \leq .04$, respectively), number of years of reading mammograms ($P \leq .03$ and $P \leq .04$, respectively), and number of mammographic readings per year ($P \leq .0001$ and $P \leq .0001$, respectively). Location sensitivity was also positively correlated with the number of mammographic readings per year ($P \leq .04$) (Table 3).

JAFROC, location sensitivity, and specificity for the three subgroups of the number of mammographic readings per year are summarized in Table 4, with significant findings summarized in Table 5. In particular, for readers who read fewer than 1000 mammograms per year, the number of years since qualification as a radiologist and the number of years of reading mammograms were inversely related to JAFROC values ($P \leq .004$ and $P \leq .002$, respectively). On the other hand, for individuals for whom the annual number of mammographic readings was greater than 5000, JAFROC values and specificity were positively related to the number of years since qualification as a radiologist ($P \leq .01$ and $P \leq .01$, respectively), along with the number of years of reading mammograms ($P \leq .002$ and $P \leq .03$, respectively) and total number of hours per week of reading mammograms ($P \leq .003$ and $P \leq .003$, respectively). For readers with an annual volume between 1000 and 5000 mammographic readings, JAFROC scores were significantly related only to number of mammographic readings per year ($P \leq .03$).

For readers with an annual volume of fewer than 1000 mammographic readings per year, multiple regression analysis suggested that a combination of the number of years of reading mammograms and the total number of

Table 5

Correlation Analysis of JAFROC, Location Sensitivity, and Specificity Values for Specific Radiologist Groupings

Parameter and Annual Volume of Readings	JAFROC		Location Sensitivity		Specificity	
	r Value	*P* Value	*r* Value	*P* Value	*r* Value	*P* Value
Age						
<1000 readings	−0.47	≤.06	−0.36	≤.15	−0.18	≤.47
1000–5000 readings	0.01	≤.92	−0.11	≤.39	0.07	≤.56
>5000 readings	0.11	≤.56	−0.19	≤.25	0.02	≤.22
No. of years since qualification as a radiologist						
<1000 readings	−0.65	≤.004*	−0.22	≤.18	−0.41	≤.09
1000–5000 readings	0.24	≤.059	0.08	≤.51	0.2	≤.12
>5000 readings	0.40	≤.01*	−0.06	≤.72	0.38	≤.01*
No. of years of reading mammograms						
<1000 readings	−0.69	≤.002*	−0.33	≤.18	−0.38	≤.12
1000–5000 readings	0.01	≤.88	0.01	≤.99	0.11	≤.37
>5000 readings	0.47	≤.002*	0.01	≤.94	0.33	≤.03*
Total no. of mammographic readings per year						
<1000 readings	0.06	≤.79	0.04	≤.85	−0.18	≤.49
1000–5000 readings	0.29	≤.03*	0.21	≤.09	0.04	≤.73
>5000 readings	0.06	≤.68	0.07	≤.63	0.35	≤.03*
No. of hours per week of reading mammograms						
<1000 readings	−0.3	≤.24	0.07	≤.78	−0.33	≤.18
1000–5000 readings	−0.10	≤.4	0.11	≤.37	−0.10	≤.43
>5000 readings	0.46	≤.003*	0.09	≤.56	0.46	≤.003*

Note.—Annual volume of readings and readings refer to annual number of mammographic readings.

* Findings with a significant difference.

hours per week of reading mammograms were more accurately predictive of JAFROC figures of merit than was either variable alone. The weights for the predictors were negative, with R^2 of 72% (F = 16.91, df = 2,13, $P \le .001$). The equation was JAFROC = 0.863 − ($Y \cdot 0.013$) − ($H \cdot 0.007$), where Y is number of years of reading mammograms and H is the number of hours per week of reading mammograms.

For readers with an annual volume of more than 5000 mammographic readings per year, the same combination of predictors was significant, R^2 of 33% (F = 8.34, df = 2,34, $P \le .001$). However, in contrast to readers with an annual volume of fewer than 1000 mammographic readings per year, the weights for the predictors were positive. The equation was JAFROC = 0.773 + ($Y \cdot 0.004$) + ($H \cdot 0.002$).

For readers with an annual volume between 1000 and 5000, only the number of mammographic readings per year emerged as a significant positive predictor (F = 5.35, df = 1,58, $P \le .024$), with a comparatively low R^2 of 8.2%. The equation was JAFROC = 0.728 + ($0.00002959 \cdot M$), where M is the number of mammographic readings per year.

Note that all constants are significantly different from zero ($P < .001$).

Discussion

The minimum annual volume of mammographic readings per year is used as a desirable level of experience in many world regions; however, this minimum value can vary substantially: 960 readings during a 2-year period in the United States (12), 2000 readings in Australia (13), 2000 readings in Canada (but 2500 readings in the province of British Columbia) (11), and 5000 readings in the United Kingdom (14). While these variations are substantial, evidence justifying such global differences is scant. In Australia, this number is based on good evidence from other researchers (11), and a recent article by our group (10) demonstrated that those radiologists with more than 2000 mammographic readings per year outperformed those with less than 1000 mammographic readings per year, with annual reading volumes of more than 5000 mammographic readings per year not necessarily yielding any additional advantage. It is important that the number is not seen as a guarantee of minimum quality, as the heterogeneity of reader characteristics and experience, even when minimum reading values are stated, is considerable, and it is possible that a reader who has appropriate experience but who has a current annual volume of less than 1000 mammographic readings per year can outperform a reader with an annual volume of more than 5000 mammographic readings per year. It is therefore necessary to explore, at a greater level than has been done previously, other key experiential factors that can affect performance, even when radiologists are grouped according to their reading volume per year.

The most obvious observation from the current work is that radiologists who read fewer than 1000 mammograms per year appear to have lower performance scores than those who read more than 1000 mammograms per year. This finding is in agreement with data in previous reports (11); however, this study has shown that, whatever the characteristics or experiences of radiologists, no factor appears to compensate for a low level of an annual volume of readings. Within individual groups of radiologists classified on the basis of the volume of readings, some interesting correlations have been demonstrated. First, readers with an annual reading volume of less than 1000 mammographic readings demonstrate reduced performance with increased years of reading mammograms. This inverse relationship between performance and years of experience has been observed to occur in other domains, such as medical general practice (19), auditor evaluations (20), chess (21), music (22–24), and different types of sports (25,26). Clearly this result has important implications for radiology, as a reasonable assumption that a low volume of readings can be compensated for by years of experience is unfounded.

On the other hand, readers with an annual volume of mammographic readings of more than 5000 demonstrated positive correlations between performance and the number of years since qualification as a radiologist, the number of years of reading mammograms, and the number of hours each week of reading mammograms. This is in direct contrast with the findings for readers with low volumes of mammographic readings and shows that, for individuals grouped as those with more than 5000 mammographic readings per year and already performing at a high level, they get even better at this activity with more time. Further examination of the results shows that this finding is not linked with improved detection of cancer, because correlation values are close to zero for this parameter; instead, this finding is linked to the increased ability to recognize normal images (specificity) (27,28). This result implies that the true discriminating factor that separates individuals performing at the highest levels from others is the ability to recognize what is normal. In screening mammography, this factor clearly has important implications for clinicians and policymakers who are encouraging a reduction in recall rates for further assessment after initial screening readings (while not affecting cancer detection).

We acknowledge that there were limitations in this work. First, the number of cases assessed was relatively small, and the case mix was not typical of a screening environment. Second, it would have been preferable to have the decision-recording software (Ziltron Technologies) loaded onto the workstations, but both picture archiving and communication system companies were unwilling to have this tool integrated with the picture archiving and communication system software. Third, although the r values were larger when individuals were allocated to volume-dependent groupings (up to -0.69) compared with the single overall grouping, these values generally remained low. Fourth, we cannot rule out that radiologists who performed less well avoid interpreting mammographic images and that low volume is not necessarily driving the outcome of low performance. Finally, comparisons with prior images were not included, and this lack comparison could have influenced results.

In conclusion, the current work demonstrates that reading volumes are associated with reader performance variations; however, examination of individual readers grouped according to different levels of reading volumes demonstrates other key agents that affect performance. The ability to recognize normal images is an important discriminatory function in individuals who read high volumes of mammograms.

Disclosures of Conflicts of Interest: M.A.R. No relevant conflicts of interest to disclose. **W.B.L.** No relevant conflicts of interest to disclose. **R.M.B.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: paid for MRI lectures to radiology trainees by RANZCR and institution received payment for development of a physics module from RSNA/AAPM. Other relationships: none to disclose. **E.A.R.** No relevant conflicts of interest to disclose. **M.W.P.** No relevant conflicts of interest to disclose. **W.M.R.** No relevant conflicts of interest to disclose. **R.C.H.** No relevant conflicts of interest to disclose. **D.A.B.** No relevant conflicts of interest to disclose. **P.C.B.** No relevant conflicts of interest to disclose.

References

1. Poortmans P, Aznar M, Bartelink H. Quality indicators for breast cancer: revisiting historical evidence in the context of technology changes. Semin Radiat Oncol 2012;22(1):29–39.
2. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137(5 pt 1): 347–360.
3. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample. Arch Intern Med 1996;156(2):209–213.
4. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. J Natl Cancer Inst 2005;97(5):358–367.
5. Barlow WE, Chi C, Carney PA, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. J Natl Cancer Inst 2004;96(24):1840–1850.
6. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. Radiology 2002;224(3): 861–869.
7. Brem RF, Baum J, Lechner M, et al. Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial. AJR Am J Roentgenol 2003;181(3):687–693.
8. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol 1978;13(3):175–181.
9. Miglioretti DL, Smith-Bindman R, Abraham L, et al. Radiologist characteristics associated with interpretive performance of diagnostic mammography. J Natl Cancer Inst 2007;99(24):1854–1863.
10. Reed WM, Lee WB, Cawson JN, Brennan PC. Malignancy detection in digital mammograms: important reader characteristics and required case numbers. Acad Radiol 2010;17(11):1409–1413.
11. Kan L, Olivotto IA, Warren Burhenne LJ, Sickles EA, Coldman AJ. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. Radiology 2000;215(2):563–567.
12. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. Radiology 2009;253(3):641–651.
13. U.S. Department of Health and Human Services. An overview of the final regulations implementing the Mammography Quality Standards Act of 1992. Rockville, Md: U.S. Department of Health and Human Services, 1997; 16–19.
14. National Accreditation Committee. National Program for the Early Detection of Breast Cancer—National Accreditation Requirements. Canberra, Australia: Commonwealth Department of Human Services and Health, 1994.
15. National Health Service Breast Screening Radiologists Quality Assurance Committee. Quality assurance guidelines for radiologists. National Health Service Breast Screening

Programme publication no. 15. Sheffield, England: NHSBSP Publications, 1997.

16. Obuchowski NA. ROC analysis. AJR Am J Roentgenol 2005;184(2):364–372.

17. Chakraborty DP, Winter LH. Free-response methodology: alternate analysis and a new observer-performance experiment. Radiology 1990;174(3 pt 1):873–881.

18. Chakraborty DP, Breatnach ES, Yester MV, Soto B, Barnes GT, Fraser RG. Digital and conventional chest imaging: a modified ROC study of observer performance using simulated nodules. Radiology 1986;158(1): 35–39.

19. Butterworth JS, Reppert EH. Auscultatory acumen in the general medical population. JAMA 1960;174(1):32–34.

20. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 2004;79(10 suppl):S70–S81.

21. Lehmann AC, Ericsson KA. Music performance without preparation: structure and acquisition of expert sight-reading. Psychomusicology 1996;15(1-2):1–29.

22. Charness N, Krampe R, Mayr U. The role of practice and coaching in entrepreneurial skill domains: an international comparison of life-span chess skill acquisition. In: Ericsson KA, ed. The road to excellence: the acquisition of expert performance in the arts and sciences, sports, and games. Mahwah, NJ: Lawrence Erlbaum Associates, 1996; 51–80.

23. Sloboda JA. The acquisition of musical performance expertise: deconstructing the"talent" account of individual differences in musical expressivity. In: Ericsson KA, ed. The road to excellence: the acquisition of expert performance in the arts and sciences, sports, and games. Mahwah, NJ: Lawrence Erlbaum Associates, 1996; 107–126.

24. Starkes JL, Deakin J, Allard F, Hodges NJ, Hayes A. Deliberate practice in sports: what is it anyway? In: Ericsson KA, ed. The road to excellence: the acquisition of expert performance in the arts and sciences, sports, and games. Mahwah, NJ: Lawrence Erlbaum Associates, 1996; 81–106.

25. Helsen WF, Starkes JL, Hodges NJ. Team sports and the theory of deliberate practice. J Sport Exerc Psychol 1998;20(1):12–34.

26. Ward P, Hodges NJ, Williams AM, Starkes JL. Deliberate practice and expert performance. In: Williams AM, Hodges NJ, eds. Skill acquisition in sport. London, England: Routhledge, 2004; 231–258.

27. Elmore JG, Miglioretti DL, Reisch LM, et al. Screening mammograms by community radiologists: variability in false-positive rates. J Natl Cancer Inst 2002;94(18):1373–1380.

28. Alberdi RZ, Llanes AB, Ortega RA, et al. Effect of radiologist experience on the risk of false-positive results in breast cancer screening programs. Eur Radiol 2011;21(10): 2083–2090.