Published in final edited form as:

AJR Am J Roentgenol. 2012 September ; 199(3): 695–702. doi:10.2214/AJR.11.7402.

Radiologists' ability to accurately estimate and compare their own interpretative mammography performance to their peers

Andrea J. Cook, PhD^{1,2}, Joann G. Elmore, MD, MPH³, Weiwei Zhu, MS¹, Sara L. Jackson, MD, MPH³, Patricia A. Carney, PhD⁴, Chris Flowers, MD⁵, Tracy Onega, PhD⁶, Berta Geller, EdD⁷, Robert D. Rosenberg, MD⁸, and Diana L. Miglioretti, PhD^{1,2}

¹Biostatistics Unit, Group Health Research Institute, Seattle, WA 98101

²Department of Biostatistics, University of Washington, Seattle, WA 98195

³Department of Internal Medicine, University of Washington, Seattle, WA 98195

⁴Departments of Family Medicine and Public Health and Preventive Medicine, Oregon Health & Science University, Portland, OR 97239

⁵Department of Diagnostic Imaging, Moffitt Cancer Center, Tampa, FL 33612

⁶Department of Community and Family Medicine, Dartmouth Medical School, Lebanon, NH 03755

⁷Family Practice and Radiology, University of Vermont, Burlington, VT 05405

⁸Department of Radiology, University of New Mexico, Albuquerque, NM 87131

Abstract

Objective—To determine if U.S. radiologists accurately estimate their own interpretive performance of screening mammography and how they compare their performance to their peers'.

Materials and Methods—174 radiologists from six Breast Cancer Surveillance Consortium (BCSC) registries completed a mailed survey between 2005 and 2006. Radiologists' estimated and actual recall, false positive, and cancer detection rates and positive predictive value of biopsy recommendation (PPV2) for screening mammography were compared. Radiologists' ratings of their performance as lower, similar, or higher than their peers were compared to their actual performance. Associations with radiologist characteristics were estimated using weighted generalized linear models. The study was approved by the institutional review boards of the participating sites, informed consent was obtained from radiologists, and procedures were HIPAA compliant.

Results—While most radiologists accurately estimated their cancer detection and recall rates (74% and 78% of radiologists), fewer accurately estimated their false positive rate and PPV2 (19% and 26%). Radiologists reported having similar (43%) or lower (31%) recall rates and similar (52%) or lower (33%) false positive rates compared to their peers, and similar (72%) or higher (23%) cancer detection rates and similar (72%) or higher (38%) PPV2. Estimation accuracy did not differ by radiologists' characteristics except radiologists who interpret 1,000 mammograms annually were less accurate at estimating their recall rates.

Conclusion—Radiologists perceive their performance to be better than it actually is and at least as good as their peers. Radiologists have particular difficulty estimating their false positive rates and PPV2.

INTRODUCTION

There is a movement to assess physicians' clinical performance in everyday practice to allow identification of poorly performing physicians and to guide quality improvement and professional development [1]. However, little is known about how well physicians understand their own performance measures. A recent review found that thirteen out of twenty studies demonstrated little, no, or an inverse relationship between self-assessment measures and physician indicators of clinical performance [2]. However, the physician populations evaluated in these studies were not concurrently receiving information on their performance to give them the ability to accurately self-assess themselves.

Radiologists who interpret mammography are an example of a unique specialty that has received standardized performance information in the U.S. since the 1990's through the Mammography Quality Standards Act (MQSA) requirement that mammography facilities "collect and review outcome data for all mammograms performed" [3]. The extent to which radiologists review and retain information provided by the audit reports is unknown. A concern is that if radiologists are not accurate in their perceptions of their performance, they will not know if, and in which areas, their performance needs improvement. Criteria have been established in screening mammography for identifying low performing radiologists who might benefit from educational interventions [4].

A 2001 survey of U.S. radiologists, from 3 mammography centers, found that they overestimated their rates of recommending further evaluation after screening mammograms [5]. We have extended this prior research by assessing a larger number of radiologists from 3 more geographic sites than the previous report after a five year interval of performance feedback. We also evaluate interpretive performance for a broader array of performance measures, including false-positive and cancer detection rates. In this report we also assess radiologists' perceptions of how their interpretative performance compares to the performance of other U.S. radiologists. Peer comparison feedback has resulted in improvements in physician performance in other areas [6, 7] and comparisons to others may be important to fully appreciate any performance gaps. In this study we tested the hypothesis that radiologists would not be able to accurately estimate their performance, but would be better at rating how their performance compares to their peers. We also hypothesized that some radiologist characteristics, including receipt of audit reports, clinical experience, and fellowship training, would increase radiologists' accuracy at estimating their own interpretive performance.

MATERIALS AND METHODS

Six mammography registries from the National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC; <http://breastscreening.cancer.gov>) contributed data for this study: San Francisco (SF), North Carolina (NC), New Mexico (NM), New Hampshire (NH), Vermont (VT), and Western Washington (WA). These registries collect patient demographic and clinical information each time a woman receives a mammography examination at a participating facility. This information is linked to regional cancer registries and pathology databases to determine cancer outcomes. Data from the registries were pooled at the BCSC Statistical Coordinating Center (SCC) for analysis.

Radiologists who interpreted mammograms at a BCSC facility between January 2005 and December 2006 were invited to participate in a mailed survey in 2005 (NH, VT, and WA) or

2006 (SF, NC, NM), using survey methods previously described ([8], Survey: http://breastscreening.cancer.gov/collaborations/favor_ii_mammography_practice_survey.pdf). Included in the analysis were mammograms indicated to be routine screening and interpreted by a surveyed radiologist between January 1, 2000 and December 31, 2005 or December 31, 2006, depending upon when the respective site administered the survey. This date restriction was made because radiologists were asked to estimate their screening interpretive performance since 2000 at the time of taking the survey. Of those 212 survey completers (68.6% complete rate=214/312), 38 additional radiologists were removed due to not performing any screening mammography during the study period for a total of 174 radiologists eligible for the study.

Each registry and the SCC received Institutional Review Board (IRB) approval to consent radiologists to take the survey and to link their survey data to BCSC performance data. Sites also received IRB approval for either active, or passive, consenting processes or a waiver of consent to collect women-level and mammography data, link data, and perform analytic studies. All procedures are compliant with Health Insurance Portability and Accountability Act (HIPAA) and all registries and the SCC have received a Federal Certificate of Confidentiality and other protection for the identities of women, physicians, and facilities that are subjects of this research.

For analyses comparing actual performance outcomes and radiologists' estimated performance, we restricted to the 155 radiologists (155/174= 89%) that provided an estimate to at least one of the survey questions on recall, false-positive, or cancer detection rates. These performance outcomes were based on the initial assessment of 1,082,639 screening mammograms interpreted by the 155 radiologists. Analysis of the positive predictive value of biopsy recommended (PPV2) was based on the final assessment after all imaging was performed up to 90 days after the screen and prior to the breast biopsy from 990,574 screening mammograms interpreted by 132 radiologists (132/174=76%).

Measurements

Data on radiologist characteristics were obtained from the survey, which included the following questions on screening performance: 1) "For the following questions about screening mammograms, please estimate your values since 2000 for recall rate (%), PPV2 (%), False-positive Rate (%), and Cancer Detection Rate (per 1000)" and 2) "How do you think your current screening performance compares to others in the U.S.?" with a likert scale for response of "Much Lower", "Lower", "Same", "Higher", or "Much Higher". The following definitions were provided in the survey: "recall rate is % of all screens with a positive assessment leading to immediate additional work-up"; "positive predictive value of biopsy recommended (PPV2) is % of all screens with biopsy or surgical consultation recommended that resulted in cancer"; "false-positive rate is % of all screens interpreted as positive and no cancer is present"; and "cancer detection rate is # of cancers detected by mammography per 1000 screens". We define the estimated performance values as the radiologist's *perceived performance*. We categorized radiologist's *perceived relative performance* compared to others in the US into 3 categories: 1) lower (much lower or lower), 2) same, and 3) higher (higher or much higher).

The survey included demographic, experience, and clinical practice characteristics in the prior year, including age, sex, affiliation with an academic institution, fellowship training, years of mammography experience, hours working in breast imaging per week, number of continuing medical education (CME) hours in breast imaging in the last 3 years, self-reported number of mammograms interpreted in the prior year and if the radiologist received audit reports with information on their own interpretive performance (yes/no). We also included responses to questions evaluating the radiologist's perspective of their confidence

in use of numbers and statistics and frequency of using numbers and statistics when discussing a positive mammogram with a patient, which were both measured on a 5-point likert scale. We defined *numeric confidence* as reporting being confident, or very confident, in both interpreting medical literature and audit reports. We categorized how often a radiologist reports using numbers and statistics when discussing a positive mammogram with a patient into three categories: 1) never or rarely, 2) sometimes, and 3) often or always.

We defined a mammogram as being associated with a cancer if the woman was diagnosed with invasive breast cancer or ductal carcinoma *in situ* (DCIS) within one year of the mammography examination and before the next screening mammogram. We used standard definitions developed by the BCSC to define a mammogram assessment as *positive* and to calculate interpretative performance measures [9].

To evaluate how well the radiologists estimated their performance measures compared to their actual performance measures we created a binary outcome *accurate perceived performance* defined as accurate if they estimated their performance within ± 5 percentage-points of their actual performance (i.e. if actual false-positive rate is 10% then a radiologist whom provides an estimated false-positive rate between 5% and 15% would be defined as accurate) and inaccurate otherwise. The ± 5 percentage-point cut-off was chosen *a priori* and allowed for radiologists to round their estimated performance and still be defined as accurate.

Statistical Analysis

To visually assess how well radiologists were able to estimate their actual performance outcomes, we plotted perceived versus actual performance for each of the four performance outcomes. We calculated the concordance correlation coefficient [10] to summarize how well perceived performance agreed with actual performance. For cancer detection rate we conducted the analysis without six outliers that estimated cancer detection rate as greater than 25 per 1,000, because they were unduly influential.

For each performance measure, we calculated the mean perceived performance and percent of radiologists who had accurate perceived performance, evaluated their unadjusted associations with radiologist characteristics, and calculated 95% confidence intervals(CI) using linear and logistic regression, respectively. For each perceived performance question, we calculated the frequency that radiologists provided no response, and the frequency perceived performance was >5 percentage-points below actual performance, an accurate estimate, and >5 percentage points above actual performance.

We calculated the unadjusted associations between a radiologist's perceived relative performance and their perceived and actual performance. For actual performance outcomes we applied weighted regression with weights corresponding to number of mammograms interpreted by a radiologist for a given analysis population. Specifically, for actual recall and cancer detection rates, weights were the number of screening mammograms interpreted by the radiologist; for false-positive rate weights were the number of mammograms without a diagnosis of cancer; and for PPV2 weights were the number of mammograms interpreted by the radiologist as positive at the end of the imaging work-up.

All confidence intervals and p-values are two-sided based on the Wald statistic with statistically significant associations at the $P<0.05$ level. Data analyses were conducted using SAS® software, Version 9.2 (SAS institute, Cary, NC).

RESULTS

Radiologist's Perception of Performance

Of the 174 radiologists who returned the survey and had interpreted screening mammograms during the study period, the percentage providing estimates of their performance varied by the measure: recall rate 89%(155/174); false-positive rate 63%(110/174); cancer detection rate 76%(133/174); and PPV2 76%(132/174).

Most radiologists accurately estimated their recall (78%) and cancer detection (72%) rates, but only 19% and 26% accurately estimated their false-positive and PPV2 rates, respectively (Table 1). While all 155 radiologists completed the survey question estimating their recall rate, the number responding to the other measures was much lower (n=110 false-positive rate, n=127 cancer detection rate, n=132 PPV2). The most common reason for low accuracy of perceived false-positive rate was due to not providing an estimate.

Only 50% of radiologist who interpreted 1,000 mammograms annually accurately estimated recall rate compared to 73% of radiologists who interpreted 1001–2000 mammograms and 80% who interpreted >2000 mammograms (Table 1). Those who reported never, or rarely, using numbers or statistics when discussing mammography results with patients were less accurate in estimating their own cancer detection rate (66%) compared to those who sometimes (83%) or often/always (90%) use numbers or statistics when communicating with patients. No other radiologist characteristics were statistically significantly related to perceived performance accuracy.

Among all radiologists, 19.4%(30/155) underestimated and 2.6%(4/155) overestimated their actual recall rate, while 35%(53/155) overestimated their false-positive rate, 50%(57/114) overestimated PPV2, and 11%(17/155) overestimated their cancer detection rate (Table 2). Among radiologists who provided a performance measurement, 49%(53/108) overestimated false-positive rate and 13%(17/131) overestimated cancer detection rate.

Figure 1 shows the relationship between actual and perceived performance. Recall rate was estimated by radiologists with the most accuracy (concordance correlation (Corr) =0.55) compared to PPV2 (Corr=0.32), cancer detection rate (Corr=0.13, excluding 6 outliers), and false-positive rate (Corr=0.02).

Radiologist's Perception of Performance Compared with their Peers

Radiologists in general perceived their screening performance as equal, or better, relative to others (Table 3). Radiologists reported having similar (43%) or lower (31%) recall rates compared to other radiologists and similar (52%) or lower (33%) false-positive rates. For cancer detection rate, radiologists perceived having similar (72%) or higher (23%) values and similar (72%) or higher (38%) values for PPV2. Only a few radiologists viewed their cancer detection rate (5%) or PPV2 (3%) as worse than their peers.

Radiologists with low, medium, or high perceived relative performance for recall rate and PPV2 had correspondingly low, medium, or high actual and estimated performance measures (Figure 1). There was no obvious relationship with perceived relative performance and actual, or estimated, false-positive or cancer detection rates. Table 3 similarly shows radiologists who perceive their recall rate as being lower relative to their peers have lower mean actual, and perceived recall rates of 7.8%, and 6.0%, compared to 15.0%, and 13.5%, respectively, among those who perceive having a higher relative recall rate. Similar patterns were observed for PPV2 and cancer detection rate, but only for actual, and not perceived false-positive rate. Further, radiologists who perceived their recall rate as lower relative to their peers were more accurate in estimating their actual recall rate (92% accurate) compared

to those who perceive their recall rates as the same (66% accurate) or higher (83% accurate). Similar results occurred for false-positive rate and PPV2, but not for cancer detection rate in which those whom perceived having a lower cancer detection rate relative to their peers were less accurate in estimating their actual cancer detection rate (63% accurate) compared to those perceiving having the same or higher (both 75% accurate).

DISCUSSION

Radiologists reading mammograms are mandated to produce performance data (audit reports) for their MQSA certified institution, and these measures are deliberately designed to be used as a performance improvement metric. However, in our study we noted that radiologists were relatively good at estimating their recall and cancer detection rates, but most were unable to accurately estimate their false-positive rate or PPV2. Radiologists tended to underestimate their recall rate and overestimate their false-positive, cancer detection, and PPV2 rates. Many radiologists perceive themselves as having better interpretative performance than they actually do. This is an important finding, because without an accurate understanding of their performance, it is unrealistic to expect radiologists to know whether improvement is needed and which areas are most in need of improvement. Performance feedback should include both definitions of the performance measures and display results relative to national guidelines or peer performance to assist highly motivated physicians to improve if needed.

While almost all radiologists (96%) reported receiving audit reports, receipt of these data did not appear to fully inform them of their own performance on the outcome measures in this study. We had hypothesized that receipt of audit reports, clinical experience and fellowship training would all improve radiologists' accuracy at estimating their own interpretative performance, but we found minimal evidence of this relationship. Only radiologists with a higher volume of mammograms had a positive effect on accurately estimating recall rate and radiologists who more frequently used numbers or statistics when discussing mammography results with patients were more accurate in estimating their cancer detection rate. While audit report information was available to all of the study radiologists since they work at a BCSC facility, the audit data varied across sites with most sites providing information at the radiologist level, but others at the facility level only [11]. The type of information provided also varies across sites, with all reports providing recall rates and none reporting false-positive rates. An important next step would be to evaluate how audit reports are actually reviewed and considered by individual radiologists. Research on physician behavior change indicates that predisposing physicians to change requires showing them the gap between their own performance and that of national targets [12, 13]. Prior work in this area suggests that the format of audit reports may make a difference in how physicians use data to improve their clinical practice [6].

Another important finding of our study was how few radiologists were able to accurately estimate their false-positive rate. Only 62% of radiologists even attempted to provide an estimate and among those who did, only 28% provided an accurate estimate. The provided values suggest some radiologists may have confused false-positive rate and specificity when completing the survey, even though definitions were provided. However, even if we assume that radiologists who provided very high estimates of their false-positive rate (i.e., >50%) were actually providing estimates for their specificity and calculate false-positive rate from those values as 100-specificity, still only 36% accurately estimated their false-positive rate. Further, a recent study on this same population evaluating if radiologists could predict a reasonable goal for these performance measures, only 22% reported goals for false-positive rate within the range recommended by the American College of Radiology [14]. False-positive rates are much higher in the US relative to Canada and European countries with

screening programs [15] likely due in part to malpractice concerns. However, this represents an opportunity for improvement in order to minimize the negative consequences of over-treatment, anxiety and cost for women [16]. It is also possible that radiologists interpreting screening mammograms do not typically conduct the diagnostic work-ups on these same patients, thus they are not always aware of the outcome. It will be difficult to motivate radiologists to reduce their false-positive rates (while maintaining sensitivity) if they do not understand what their false-positive rates currently are, how it is calculated, or how they compare with their peers.

Only one previous study compares estimated versus actual mammography performance, but it was restricted to three geographic regions in the US and did not evaluate false-positive and cancer detection rates [5]. Our analysis was also conducted on a more recent survey, such that radiologists had 5 additional years of audit feedback about their performance, and had more cumulative years of actual performance data from the BCSC to accurately estimate their performance. Our results are not directly comparable to this early report as the statistical methods were different and the screening population was more restrictive in the previous study.

Our study has several strengths. We had a good response to our survey tool (68.6%) and we examined radiologists' perceived interpretive performance compared to their peers and also their actual mammography performance data from clinical practices in six geographically distinct regions in the US. This suggests that the findings of our study are generally applicable across the country.

One weakness of our study was the low response rate for estimating false-positive and cancer detection rates. This may indicate that participants were not comfortable estimating these two measures. We suspect that radiologists purposely skipped these questions, as all but one radiologist who did not answer these questions provided responses for the subsequent survey questions. Having variable audit data across BCSC sites is also a limitation. However, since almost all study radiologists reported receiving audit reports, radiologists could have looked up their performance measures in their BCSC audits; thus, our results could overestimate the percentage of US radiologists who can accurately describe their interpretive performance.

For most performance measures, radiologists overestimate their ability, including perceiving their screening interpretive performance as better than their peers and have particular difficulty in estimating their false-positive rate and PPV2. Given the study findings, opportunities for improving radiologists' understanding of their performance could include a standardized facility and physician audit reporting form such as a "Radiologist Report Card" for screening mammography interpretation, with clear reporting of recall rate, PPV2, false positive rate, and cancer detection rate relative to national guidelines or peer cohort. Similarly, providing a website that allows for radiologists to compare themselves to their peers in the United States and other countries may improve their ability to understand their own interpretive performance measures, and know if, and in which areas, their interpretive performance needs improvement. [17] Development of widely available CME specific to MSQA reporting of performance measures, including radiologist's individual audit data relative to peers', could also be an effective tool for self assessment, and potentially ultimately improve clinical interpretation. The routine submission of local data to the National Mammography Database, developed by the American College of Radiology, <https://nrdr.acr.org/Portal/NMD/Main/page.aspx> (accessed 12/1/11), is another powerful tool which should assist radiologists in their understanding of their own performance relative to their peers

Conclusion

Radiologists perceive their performance to be better than it actually is and at least as good as their peers. Radiologists have particular difficulty estimating their false positive rates and PPV2. Future study of strategies to improve audit feedback to and education of radiologists is warranted, but encouragement for radiologists to join the ACR National Mammography Database would answer many of these findings.

Acknowledgments

This work was supported by the National Cancer Institute and Agency for Healthcare Research and Quality (R01 CA107623), the National Cancer Institute (K05 CA104699; Breast Cancer Surveillance Consortium: U01CA63740, U01CA86076, U01CA86082, U01CA70013, U01CA69976, U01CA63731, U01CA63736 U01CA70040), the Breast Cancer Stamp Fund, and the American Cancer Society, made possible by a generous donation from the Longaberger Company's Horizon of Hope[®] Campaign (SIRGS-07-271-01, SIRGS-07-272-01, SIRGS-07-273-01, SIRGS-07-274-01, SIRGS-07-275-01, SIRGS-06-281-01). The collection of cancer data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: <http://breastscreening.cancer.gov/work/acknowledgement.html>. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. We thank the participating women, mammography facilities, and radiologists for the data they have provided for this study. A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: <http://breastscreening.cancer.gov/>.

References

1. Scott IA, Phelps G, Brand C. Assessing individual clinical performance - A primer for physicians. *Intern Med J*. 2010
2. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006; 296:1094–1102. [PubMed: 16954489]
3. US Food and Drug Administration. MQSA Final Regulations. 2010.
4. Carney PA, Sickles EA, Monsees BS, et al. Identifying minimally acceptable interpretive performance criteria for screening mammography. *Radiology*. 2010; 255:354–361. [PubMed: 20413750]
5. Fenton JJ, Egger J, Carney PA, et al. Reality check: perceived versus actual performance of community mammographers. *AJR Am J Roentgenol*. 2006; 187:42–46. [PubMed: 16794153]
6. Aiello Bowles EJ, Geller BM. Best ways to provide feedback to radiologists on mammography performance. *AJR Am J Roentgenol*. 2009; 193:157–164. [PubMed: 19542408]
7. Cammisa C, Partridge G, Ardans C, Buehrer K, Chapman B, Beckman H. Engaging Physicians in Change: Results of a Safety Net Quality Improvement Program to Reduce Overuse. *Am J Med Qual*. 2010
8. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology*. 2009; 253:641–651. [PubMed: 19864507]
9. Breast Cancer Surveillance Consortium. BCSC Glossary of Terms. 2009.
10. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45:255–268. [PubMed: 2720055]
11. Elmore JG, Aiello Bowles EJ, Geller B, et al. Radiologists' attitudes and use of mammography audit reports. *Acad Radiol*. 2010; 17:752–760. [PubMed: 20457418]
12. Laidley TL, Braddock IC. Role of Adult Learning Theory in Evaluating and Designing Strategies for Teaching Residents in Ambulatory Settings. *Adv Health Sci Educ Theory Pract*. 2000; 5:43–54. [PubMed: 12386475]
13. Speck M. Best Practice in Professional Development for Sustained Educational Change. *ERS Spectrum*. 1996; 14:33–41.
14. Jackson, SL.; Cook, AJ.; Miglioretti, DL., et al. Academic Radiology. 2011. Are radiologists' goals for mammography accuracy consistent with published recommendations?. In Press

15. Fletcher SW, Elmore JG. False-positive mammograms--can the USA learn from Europe? *Lancet*. 2005; 365:7–8. [PubMed: 15639661]
16. Nelson HD, Tyne K, Naik A, et al. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2009; 151:727–737. W237–742. [PubMed: 19920273]
17. Geller BM, Ichikawa L, Miglioretti DL, Eastman D. Web-based mammography audit feedback. *Am J Roentgenol*. 2011 In Press.

Figure 1.

Scatter plot of radiologist's actual compared to their estimated performance values for (a) recall rate, (b) false-positive rate, (c) cancer detection rate, and (d) PPV2 categorized by their self-reported relative perception of their performance value compared to radiologists in the US (Higher, Same, or Lower).

Table 1

Radiologist characteristics by radiologists' perceived performance and the percent of radiologists whose perceived performance is within $\pm 5\%$ of actual performance.

Radiologists Characteristics	Self Reported Perceived Performance						Accurate Perceived Performance*					
	Recall Rate			FPR			CDR			PPV2		
	N	mean (95%CI)	N	mean (95%CI)	N	mean (95%CI)	N	mean (95%CI)	Recall Rate	FPR	CDR	PPV2
Total	155	9.5 (8.7, 10.2)	110	33.4 (27.0, 39.9)	127	5.3 (5.3, 4.8)	132	31.0 (27.8, 32.2)	78	19	75	26
Age at Survey												
34–44	44	11.0 (9.6, 12.3)	34	31.8 (20.5, 43.1)	33	5.2 (4.2, 6.1)	38	26.8 (20.6, 33.0)	68	20	71	11
45–54	56	9.8 (8.6, 11.1)	38	42.4 (30.9, 53.9)	47	5.4 (4.8, 6.1)	49	31.9 (27.6, 36.3)	80	13	71	30
55	55	7.8 (6.5, 9.2)	38	26.0 (15.0, 37.0)	47	5.3 (4.3, 6.2)	45	33.7 (27.3, 40.0)	84	25	81	34
Sex												
Male	106	8.6 (7.7, 9.5)	80	36.3 (28.6, 44.1)	87	5.0 (4.4, 5.6)	91	31.7 (27.7, 35.6)	77	21	74	22
Female	49	11.2 (9.9, 12.6)	30	25.8 (14.0, 37.5)	40	6.0 (5.2, 6.9)	41	29.6 (24.0, 35.2)	80	16	77	33
Affiliation with Academic Medical Center												
No	124	9.2 (8.3, 10.0)	88	35.2 (27.8, 42.6)	102	5.2 (4.7, 5.7)	107	30.8 (27.4, 34.3)	78	21	73	25
Adjunct	15	10.6 (6.8, 14.4)	11	20.6 (5.9, 35.3)	11	6.4 (3.3, 9.5)	11	40.7 (21.4, 59.9)	67	13	64	18
Primary	13	10.8 (8.3, 13.3)	10	27.8 (1.7, 53.8)	12	4.7 (4.0, 5.4)	12	26.7 (20.8, 32.5)	92	15	100	33
Fellowship training												
No	141	9.5 (8.7, 10.3)	99	34.5 (27.6, 41.3)	115	5.3 (4.8, 5.8)	120	31.1 (27.8, 34.3)	77	19	73	27
Yes	14	9.0 (7.3, 10.7)	11	24.2 (2.1, 46.4)	12	5.3 (3.8, 6.8)	12	30.8 (15.4, 46.2)	86	21	86	9
Years of mammography interpretation												
<10	31	10.9 (9.3, 12.6)	24	46.2 (31.0, 61.4)	26	5.1 (4.1, 6.1)	28	30.5 (22.8, 38.2)	71	13	70	23
10–19	56	9.8 (8.7, 10.8)	40	36.5 (25.4, 47.5)	45	5.5 (4.7, 6.4)	48	28.1 (23.2, 33.1)	77	16	70	18
20	67	8.5 (7.2, 9.9)	45	24.5 (15.2, 33.8)	55	5.2 (4.5, 6.0)	55	33.8 (28.6, 38.9)	82	24	80	36
Hours working breast imaging												
0–30	106	9.6 (8.6, 10.5)	75	37.4 (29.3, 45.5)	88	5.3 (4.7, 5.9)	93	32.5 (28.6, 36.3)	80	19	74	25
31–40	18	8.8 (7.1, 10.4)	13	20.9 (4.4, 37.4)	14	6.1 (4.8, 7.4)	13	27.3 (15.2, 39.3)	78	22	72	18
>40	22	9.7 (8.2, 11.3)	15	30.2 (10.6, 49.8)	17	5.3 (3.7, 6.8)	19	23.5 (17.3, 29.8)	73	18	70	35
Number of CME hours in last 3 yrs												
15 hr minimum	36	10.4 (8.4, 12.5)	26	28.8 (16.0, 41.6)	28	5.0 (4.1, 5.8)	29	33.5 (24.7, 42.3)	69	22	68	14

Radiologists Characteristics	Recall Rate				Self Reported Perceived Performance				Accurate Perceived Performance*			
	N	mean (95%CI)	N	mean (95%CI)	FPR	N	mean (95%CI)	CDR	PPV2	Recall Rate	FPR	CDR
>15 hrs – 30 hrs	70	9.6 (8.5, 10.7)	47	36.2 (26.1, 46.3)		58	5.7 (4.8, 6.5)		31.2 (26.5, 35.9)	80	16	72
30 hrs	48	8.2 (7.2, 9.3)	37	33.3 (21.3, 45.2)		41	5.0 (4.4, 5.7)		29.2 (24.4, 34.0)	83	23	85
Average # of mammograms interpreted per year												
1000	12	14.5 (9.7, 19.2)	8	39.9 (7.9, 71.9)		7	8.2 (5.5, 10.8)		33.5 (15.4, 51.6)	50	8	50
1001–2000	55	9.8 (8.5, 11.2)	39	27.3 (17.5, 37.0)		50	5.2 (4.4, 6.0)		30.1 (24.3, 35.9)	73	20	76
> 2000	80	8.5 (7.7, 9.3)	59	38.4 (28.9, 47.8)		64	5.2 (4.5, 5.8)		31.6 (27.3, 35.8)	85	19	76
Numeric Confidence**												
No	57	10.6 (9.3, 12.0)	37	29.3 (18.9, 39.8)		48	5.6 (4.9, 6.4)		35.2 (29.3, 41.1)	82	18	77
Yes	98	8.8 (7.8, 9.7)	73	35.5 (27.3, 43.8)		79	5.1 (4.5, 5.7)		28.8 (25.0, 32.5)	76	20	73
Use of Numbers and Statistics***												
Never/rarely	78	9.4 (8.3, 10.5)	54	27.6 (19.2, 36.1)		60	5.6 (4.8, 6.4)		30.6 (25.6, 35.6)	82	27	66
Sometimes	38	9.9 (8.4, 11.3)	34	36.8 (24.7, 48.9)		35	5.6 (4.7, 6.5)		33.2 (26.2, 40.2)	74	13	83
Often/always	32	8.7 (7.3, 10.2)	19	43.2 (24.3, 62.1)		29	4.4 (3.8, 5.0)		31.0 (25.7, 36.3)	81	13	90
Receive audit reports												
No	6	11.3 (5.7, 17.0)	3	53.3 (–47.1, 153.7)		4	6.3 (2.3, 10.2)		22.0 (–26.9, 70.9)	50	17	50
Yes	146	9.3 (8.5, 10.0)	105	32.6 (26.0, 39.1)		121	5.3 (4.8, 5.8)		30.8 (27.6, 34.1)	79	20	75

Abbreviations: FPR, False-positive Rate; CDR, Cancer Detection Rate; PPV2, Positive Predictive Value; CI, Confidence Interval

* Accurate perceived performance defined as perceived performance within +/- 5% of actual performance measure

** Confident or very confident in both interpreting medical literature and audit reports.

*** Frequency of use of numbers and statistics when discussing mammography results to patients

BOLD indicates statistical significance at the 0.05 level.

Table 2

Accuracy in the perceived performance measure compared to actual performance

PERCEPTION OF PERFORMANCE COMPARED WITH ACTUAL					
Perception of performance compared with Actual	Recall	FPR	CDR	PPV2	
	N (%)	N (%)	N (%)	N (%)	N (%)
Missing	0 (0.0)	42 (28.0)	22 (14.4)	0 (0.0)	0 (0.0)
Underestimate ¹	30 (19.4)	25 (16.7)	5 (3.3)	28 (24.6)	28 (24.6)
Accurate ²	121 (78.1)	30 (20.0)	109 (71.2)	29 (25.4)	29 (25.4)
Overestimate ³	4 (2.6)	53 (35.3)	17 (11.1)	57 (50.0)	57 (50.0)

Abbreviations: FPR, False-positive Rate; CDR, Cancer Detection Rate; PPV2, Positive Predictive Value

¹ Underestimate performance defined as perceived performance < 5% of actual performance measure.
² Accurate performance defined as perceived performance within +/- 5% of actual performance measure.
³ Overestimate performance defined as perceived performance > 5% of actual performance measure.
All percents are row percents.

Estimated and actual performance by how a radiologist views their performance in relationship to their peer radiologists in the US

Table 3

How does your screening performance compare to others in the US?	Perceived Performance		Actual Performance		Accurate Perceived Performance ***	
	Mean (95% CI)		Mean (95% CI)		% Accurate	
Recall rate per 100 *						
Much lower/lower	6.0 (5.2, 6.9)		7.8 (6.8, 8.8)		92%	
Same	9.5 (8.3, 10.6)		11.9 (10.9, 12.9)		66%	
Higher/much higher	13.5 (12.3, 14.8)		15.0 (13.5, 16.5)		83%	
False-positive rate per 100 *						
Much lower/lower	31.0 (19.8, 42.2)		8.7 (7.2, 10.1)		35%	
Same	38.0 (28.9, 47.0)		11.5 (10.4, 12.5)		15%	
Higher/much higher	21.3 (3.4, 39.2)		13.5 (11.7, 15.3)		9%	
Cancer detection rate per 1000 *						
Much lower/lower	4.5 (1.8, 7.2)		4.6 (2.8, 6.3)		63%	
Same	5.1 (4.5, 5.7)		4.5 (4.0, 5.0)		75%	
Higher/much higher	6.2 (5.2, 7.1)		5.6 (4.8, 6.4)		75%	
PPV2 per 100 **						
Much lower/lower	22.1 (3.7, 40.6)		15.9 (-7.5, 39.2)		100%	
Same	26.5 (23.1, 29.9)		24.7 (21.1, 28.2)		26%	
Higher/much higher	38.9 (32.7, 45.2)		30.1 (25.1, 35.0)		19%	

* Recall, false-positive and Cancer detection rate use the initial interpretation study population: 155 radiologists; 1,082,639 mammograms, 114,238 total positive assessments, 5810 cancers which 4927 had positive assessments, and 1,076,828 non-cancers with 967,517 negative assessments.

** PPV2 uses the final interpretation study population: 132 radiologists; 990,574 mammograms, 5,504 cancers which 3,646 had final positive assessments and 985,059 non-cancers with 881,918 final negative assessments.

Recall rate includes 155 perceived and actual measurements; False-positive rate includes 108 perceived and 150 actual measurements; Cancer detection rate includes 125 perceived and 147 actual measurements; PPV2 includes 129 perceived and 114 actual measurements

*** Accurate perceived performance defined as perceived performance within +/- 5% of actual performance measure. Missing is assumed inaccurate.