



HHS Public Access

Author manuscript

AJR Am J Roentgenol. Author manuscript; available in PMC 2015 June 02.

Published in final edited form as:

AJR Am J Roentgenol. 2010 September ; 195(3): 782–785. doi:10.2214/AJR.09.2955.

Positive Predictive Value of Mammography: Comparison of Interpretations of Screening and Diagnostic Images by the Same Radiologist and by Different Radiologists

Jacqueline R. Halladay¹, Bonnie C. Yankaskas², J. Michael Bowling³, and Camille Alexander⁴

Jacqueline R. Halladay: Jacqueline_halladay@med.unc.edu

¹Department of Family Medicine, University of North Carolina School of Medicine, 590 Manning Dr., CB 7595, Chapel Hill, NC 27599-7595

²Department of Radiology, University of North Carolina School of Medicine, Chapel Hill, NC

³Department of Health Behavior and Health Education, University of North Carolina School of Medicine, Chapel Hill, NC

⁴Centra Health, Lynchburg, VA

Abstract

OBJECTIVE—The purpose of this study was to evaluate whether the positive predictive value (PPV) after a recommendation for biopsy differs when one as opposed to more than one radiologist performs the workup after abnormal findings are discovered at screening mammography.

MATERIALS AND METHODS—Using data in a mammography registry for the years 1996–2005, we identified 6,391 diagnostic examinations with a recommendation for biopsy that were performed on a day other than the day of the screening examination. The PPV after a recommendation for biopsy was calculated for two scenarios. In the first scenario, the radiologist interpreting the diagnostic images had interpreted the screening images. In the second scenario, the radiologist read diagnostic images after another radiologist had read the screening images. We used conditional logistic regression analysis to perform within-radiologist comparisons, controlling for covariates known to be associated with PPV after a recommendation for biopsy.

RESULTS—Of the screening examinations with positive findings, 2,335 (36.5%) were scenario 1, and 4,056 (63.5%) were scenario 2. We found no difference between the two scenarios with respect to PPV after a recommendation for biopsy when we controlled for age, breast density, family history of breast cancer, history of breast procedures, time since last mammogram, use of ultrasound at any point in the workup after abnormal results of screening mammography, and interval in days between the screening and diagnostic studies.

CONCLUSION—Who interprets the follow-up images after screening mammograms show abnormal findings does not appear to be an important factor influencing the wide variability in PPV among radiologists.

Keywords

breast cancer screening; diagnostic accuracy; mammography; positive predictive value

Breast cancer is second only to lung cancer as a leading cause of death of cancer among women [1]. However, the prognosis for breast cancer patients has improved, as evidenced by a reduction in mortality rates since 1989 [2], presumably because of a combination of high levels of participation in screening mammography [3], early detection, and improvements in treatment [1]. Despite this good news, there remains a wide variability in the reported accuracy of mammography and biopsy among radiologists [4, 5]. Such concerns fueled national efforts to improve breast cancer screening and led to the passage in 1993 of the Mammography Quality Standards Act, which, in addition to setting standards for the technical and training aspects of mammography, requires that systems be in place for tracking positive mammographic findings and that radiologists interpret a minimum number of mammograms in a 2-year period [1]. Enactment of the Mammography Quality Standards Act has resulted in greater consistency in some aspects of quality, such as radiation exposure and image quality among facilities [6], but wide variability in the reported accuracy of mammography and biopsy persists [4, 5].

Variation in mammographic interpretation includes differences in recall rates, false-positive rates, false-negative rates, and predictive values of positive and negative test results [7, 8]. The published false-positive rates range from 1.5% to 24.1% [8], recall rates from 1.8% to 26.2% [7], and positive predictive value (PPV) in measurement of the probability of cancer after a recommendation for biopsy from 4.3% to 52.4% [9]. These wide variations have been attributed to differences in the populations of patients undergoing screening, processes involved in obtaining and reading mammograms, characteristics of individual radiologists, and differences in health care systems [10]. Radiologists with lower false-positive rates read larger numbers of mammograms, are older and further from medical school graduation, are men, and read a higher percentage of screening than diagnostic mammograms [4, 8, 11]. It has been suggested that most of the variability among radiologists is due to factors that have not been previously studied [8]. It is not possible to determine which features are most responsible for the variation in mammographic interpretation. Investigators are trying to understand the sources of variability in mammographic interpretation to develop more effective screening programs with lower false-positive rates but without substantially lower cancer detection rates [10].

One source of variation may have to do with whether more than one radiologist is involved in the complete workup after screening studies show abnormalities, specifically whether an individual radiologist's accuracy rates differ in the following two scenarios. In the first scenario, the radiologist interprets diagnostic studies after also interpreting the screening images. In the second scenario, this radiologist reads diagnostic studies that follow screening studies interpreted by another radiologist. It is not known whether these two scenarios differ

with respect to the probability of cancer (PPV) among women referred for biopsy. Our objective was to compare PPVs after a recommendation for biopsy in the two stated scenarios; our hypothesis was that there would be no difference.

Materials and Methods

Study Sample

For the period January 1996 through December 2005 in a mammography registry, we identified all screening mammograms of all women who underwent screening mammography and were 40 years old or older at screening. Mammograms were excluded if a woman had a personal history of breast cancer or had breast implants or if the assessment from the interpretation of the mammogram was missing. We also excluded cases in which the screening and diagnostic mammograms were obtained on a single day. We made this exclusion because the images from 99% of the additional studies performed on the day of screening mammography were read by the radiologist who interpreted the screening images, preventing within-radiologist comparison in this circumstance.

During the study period, 621,752 screening mammographic examinations were performed. In all cases, follow-up for cancer status was performed for 1 year. Of all the screening mammograms, 572,004 were interpreted as normal, and 49,948 were interpreted as having positive findings (recall rate, 8%). The 8% of patients recommended for further assessment, thus having positive findings according to BI-RADS, were included in this study. We searched the registry forward 3 months from the date of screening mammography for all imaging visits that were diagnostic studies resulting from the positive finding at the screening examination. It was possible that individual women had more than one positive screening mammogram included in the study results.

The registry used was a population-based mammography registry. Prospective information is collected from the women and the radiologists at each screening mammographic examination in mammography facilities in 39 counties in North Carolina. The health questionnaire the women visiting the registry facilities complete at each breast imaging visit includes date of birth, racial and ethnic identity, education level, history of breast procedures, family and personal history of breast cancer, hormone use, and menopausal status. This questionnaire is often completed with the help of a mammographic technologist. The technologist and radiologist record the reason for the visit, the imaging study performed, the findings, breast density, assessment of the mammographic interpretation, other imaging performed, and recommendations for follow-up. The registry is a member of the National Cancer Institute National Breast Cancer Surveillance Consortium, has a certificate of confidentiality issued by the U.S. Public Health Service, and has been approved annually by the institutional review board at our medical school.

Data Sources and Definitions

Participating radiology groups submit data to the registry on a regular basis. Mammographic data are linked with pathologic data received from pathology laboratories and the state cancer registry. An outcome is classified as cancer if the pathologic diagnosis is invasive

carcinoma or ductal carcinoma in situ. Lobular carcinoma in situ is not classified as cancer. Records are linked annually with state death records. A mammographic examination is classified as screening if two-view bilateral mammograms are obtained and the radiologist classifies the visit as a screening visit. The participating radiologists use BI-RADS for coding of assessment and breast density. The categories for interpretation assessments and management are as follows: 0, need additional imaging evaluation; 1, negative; 2, benign finding; 3, probably benign finding; 4, suspicious abnormality; 5, highly suggestive of malignancy; 6, known biopsy-proven malignancy.

In our study, to eliminate the possibility of classifying a diagnostic study (mammogram obtained for workup of a breast problem or for further evaluation of an abnormal screening mammographic finding) as a screening study, any mammogram obtained less than 9 months after a previous screening mammogram was not included as a screening mammogram. We considered findings on a screening mammogram positive if the BI-RADS category was 0, 4, or 5 or 3 if a radiologist recommended immediate follow-up. If a screening mammographic finding was assessed as positive, the case was tracked for 90 days to capture the final assessment after the diagnostic workup. If the diagnostic study resulted in a recommendation for breast biopsy, the final assessment was considered positive for calculation of the PPV of the final assessment. This final assessment was based on the original mammogram plus the diagnostic mammogram, ultrasound findings, or both. All workups were classified as mammography only or mammography plus ultrasound.

Patient data included in the descriptive and conditional logistic regression analysis included patient age (categorized for the descriptive statistics 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70+ years and a continuous variable for the conditional logistic regression modeling), breast density (BI-RADS density classification: extremely dense, heterogeneously dense, scattered fibroglandular densities, and almost entirely fat), time since last mammogram (no previous mammogram, mammogram within 36 months of the index mammogram, and last mammogram more than 36 months from the index mammogram), family history of breast cancer (first-degree relative being mother, sister, or daughter; coded yes or no), self-reported breast problem (lump, discharge, or other; coded yes or no), history of breast procedure (breast biopsy or surgery; coded yes or no), and whether the woman was taking any kind of hormone at the time of the mammographic examination (yes or no). Lack of response and responses of no were combined for the analysis of self-reported personal history of breast cancer, first-degree family history of breast cancer, and history of breast problems or procedure variables. We also had information on whether computer-aided detection was used and whether a second radiologist read the screening mammogram. The data abstraction forms are available at www.unc.edu/cmr/dataCollectionForms.shtml.

If no record was found in the database or if no date of a previous mammogram was documented in the patient report, we assumed that the patient had not undergone a previous mammographic examination. The outcome of the biopsy was cancer if biopsy results showed invasive carcinoma or ductal carcinoma in situ during the follow-up interval. The follow-up interval was defined as one calendar year after the date of the index screening

mammogram or time to the next screening mammogram as long as the period was 9 months or more.

Data Analysis

We performed descriptive analyses of the distribution of the mammograms by characteristics of the women and mammogram for the two interpretative scenarios: that in which the same radiologist interpreted the screening and diagnostic mammograms (scenario 1) and that in which a radiologist interpreted the diagnostic mammogram after another radiologist had interpreted the screening mammogram (scenario 2). Conditional logistic regression models were used because this method is used to test whether an individual radiologist's PPVs after a recommendation for biopsy differ in scenarios 1 and 2 [12, 13]. With this method, radiologists who always perform the workup after positive screening results and those who perform only diagnostic work referred from others were removed from the analysis.

We calculated PPV after a recommendation for biopsy in the two scenarios controlling for covariates of age, breast density, family history of breast cancer, history of breast procedures, time since last mammogram, and use of ultrasound at any point in the workup. Risk ratio and 95% CI were computed from fixed-effect conditional logistic analyses summarizing the association between each characteristic and the probability of a true-positive report. All patient characteristics significant at the 0.05 level remained in the final multivariate models. Conditional logistic models were fit with the SAS procedure PHREG (SAS version 9.1, SAS). A significance level of 0.05 was used for all analyses.

Results

For the years 1996–2005 the registry had a total of 621,752 screening mammograms. Of these 57,480 had positive screening findings that led to 8,038 diagnostic workup assessments with positive findings. We excluded 1,647 diagnostic assessments performed the same day as screening, leaving a total sample size of 6,391. In this total sample, 2,335 diagnostic studies (36.5%) were assessed by radiologists who also read the screening mammograms (scenario 1) and 4,056 (63.5%) were read after another radiologist had interpreted the screening mammogram (scenario 2). Women could have undergone multiple examinations in the study, but only 1.3% did so. The mammograms came from 70 facilities within 36 practices and were interpreted by 217 radiologists.

The demographic characteristics of the patients and the characteristics of the imaging studies are presented in Table 1. There were no differences between the two groups regarding age, family history of breast cancer, personal history of breast procedure, self-report of a breast problem, or current use of hormones. There was also no difference in the distribution of breast density categories or in the use of ultrasound in the workups. The unadjusted PPV after a recommendation for biopsy was 26.6% in scenario 1 and 27.9% in scenario 2, not statistically different. The results of multivariate analyses also were not significantly different after adjustment for age, breast density, family history of breast cancer, history of breast procedure, interval from previous mammogram, and use of ultrasound during workup after an abnormal screening finding (Table 2).

Discussion

Wide variation in mammographic interpretation by radiologists is evident in published data concerning several performance measurements used in mammography [4, 5, 14]. One previously unmeasured source of variation is the involvement of one or of two radiologists in screening and diagnostic assessments after abnormal findings are made on screening mammograms. We studied this question with a large population from community practices in North Carolina that participate in a mammography registry and found no difference based on whether screening and diagnostic mammograms are read by the same radiologist or different radiologists. To our knowledge, this analysis has not been done previously. We made the comparison at the level of the radiologist and did not find any difference between the accuracy of individual radiologists interpreting their diagnostic studies after their own screening mammographic interpretations and their accuracy after screening interpretations had been made by another radiologist. We therefore conclude that as of yet, this factor is not one of the unmeasured factors influencing interpretive variability.

One of the main limitations of our study was that we did not know whether the radiologists who interpreted the diagnostic studies communicated with the radiologists who interpreted the screening mammograms. Thus we are unable to provide insight into the effect of such communications. Missing data also might have had an effect on the results. We did not include current use of hormones in our model partly because of the large number of missing data (9% of study subjects had data missing on current hormone use). However, in bivariate comparisons, hormone use did not seem to be a significant factor. Double reading of screening mammograms occurred in less than 3% of the screens in both scenarios, and computer-aided detection was used in 12.3% of scenario 1 cases and 12.9% of scenario 2 cases. Thus computer-aided detection or double reading should not have affected our results. Our results reflect the practice of community radiologists, potentially limiting the ability to generalize of our results to radiologists statewide.

It is reassuring that no significant effect on PPV after a recommendation for biopsy was revealed in systems that use one format or another to perform a full workup after abnormal screening results are obtained. Because of concerns about access to mammographic services due to higher demand in the face of current and expected future shortages of radiologists performing mammography [15], our findings are informative to those charged with designing efficient full-service systems of care that maintain quality.

Acknowledgments

Partially supported by grant CA70040 from the National Cancer Institute.

References

1. Committee on New Approaches to Early Detection and Diagnosis of Breast Cancer NCPB, Board on Science, Technology, and Economic Policy, Policy and Global Affairs Division. Saving women's lives, strategies for improving breast cancer detection and diagnosis. Washington, DC: National Academies Press; 2004.

2. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomized trials. *Lancet*. 2002; 359:909–919. [PubMed: 11918907]
3. Behavioral risk factor surveillance system trends data vol. 2007. Atlanta, GA: National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention; 2007. [No authors listed]
4. Elmore JG, Miglioretti DL, Reisch LM, et al. Screening mammograms by community radiologists: variability in false-positive rates. *J Natl Cancer Inst*. 2002; 94:1373–1380. [PubMed: 12237283]
5. Esserman L, Cowley H, Eberle C, et al. Improving the accuracy of mammography: volume and outcome relationships. *J Natl Cancer Inst*. 2002; 94:369–375. [PubMed: 11880475]
6. Birdwell, RL.; Wilcox, PA. *The Mammography Quality Standards Act: benefits and burdens*. Amsterdam, The Netherlands: IOS Press; 2001.
7. Barlow WE, Chi C, Carney PA, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst*. 2004; 96:1840–1850. [PubMed: 15601640]
8. Tan A, Freeman DH Jr, Goodwin JS, Freeman JL. Variation in false-positive rates of mammography reading among 1067 radiologists: a population-based assessment. *Breast Cancer Res Treat*. 2006; 100:309–318. [PubMed: 16819566]
9. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. *Radiology*. 2006; 241:55–66. [PubMed: 16990671]
10. Elmore JG, Nakano CY, Koepsell TD, Desnick LM, D’Orsi CJ, Ransohoff DF. International variation in screening mammography interpretations in community-based programs. *J Natl Cancer Inst*. 2003; 95:1384–1393. [PubMed: 13130114]
11. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. *J Natl Cancer Inst*. 2005; 97:358–367. [PubMed: 15741572]
12. Taplin SH, Ichikawa LE, Kerlikowske K, et al. Concordance of breast imaging reporting and data systems assessments and management recommendations in screening mammography. *Radiology*. 2002; 222:529–535. [PubMed: 11818624]
13. Geller BM, Ichikawa LE, Buist DS, et al. Improving the concordance of mammography assessment and management recommendations. *Radiology*. 2006; 241:67–75. [PubMed: 16990672]
14. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. *Arch Intern Med*. 1996; 156:209–213. [PubMed: 8546556]
15. Farria DM, Schmidt ME, Monees BS, et al. Professional and economic factors affecting access to mammography: a crisis today, or tomorrow? *Cancer*. 2005; 104:491–498. [PubMed: 15973693]

Comparative Patient and Imaging Characteristics for Positive Findings Interpreted by the Same Radiologist or Different Radiologists

TABLE 1

Characteristic	Total		One Radiologist		Two Radiologists	
	No.	%	No.	%	No.	%
Total	6,391		2,335	36.5	4,056	63.5
Age (y)						
40–44	908	14.2	295	12.6	613	15.1
45–49	1,026	16.1	387	16.6	639	15.8
50–54	1,033	16.2	379	16.2	654	16.1
55–59	853	13.4	325	13.9	528	13.0
60–64	728	11.4	277	11.9	451	11.1
65–69	633	9.9	238	10.2	395	9.7
70+	1,210	18.9	434	18.6	776	19.1
Breast density						
Extremely dense	355	5.6	140	6.0	215	5.3
Heterogeneously dense	2,474	38.7	889	38.1	1,585	39.1
Scattered fibroglandular densities	3,002	47.0	1,101	47.2	1,901	46.9
Almost entirely fat	183	2.9	73	3	110	2.7
Missing information	377	5.9	132	5.7	245	6.0
Time since last mammogram ^d						
No previous mammogram	854	13.4	311	13.9	543	13.4
1–35 mo	4,197	65.7	1,535	68.7	2,662	65.5
36 mo	909	14.2	329	14.7	580	14.3
Family history of breast cancer in first-degree relative						
Yes	719	11.3	258	11.1	461	11.4
No	5,672	88.8	2,077	89.0	3,595	88.6
Self-reported breast problem						
Yes	265	4.2	99	4.2	166	4.1
No	6,126	95.9	2,236	95.8	3,890	95.9
History of breast procedure						

Characteristic	Total		One Radiologist		Two Radiologists	
	No.	%	No.	%	No.	%
Yes	1,802	28.2	640	27.4	1,162	28.7
No	4,589	71.8	1,695	72.6	2,894	71.4
Workup included ultrasound						
Yes	2,496	39.1	907	38.8	1,589	39.2
No	3,895	61.0	1,428	61.2	2,467	60.8

^a 431 variables from the total had missing data.

TABLE 2

Results of Fixed-Effect Conditional Logistic Regression Analysis for Prediction of Breast Cancer in Women With Screening Findings Positive for Breast Cancer

Variable	Unadjusted Relative Risk		Adjusted Relative Risk ^a	
	Ratio	95% CI	Ratio	95% CI
Radiologist				
Same radiologist	1.0	—	1.00	—
Different radiologist	0.99	0.87–1.13	1.02	0.88–1.18
Age (y)				
40–49	1.0	—	1.00	—
50–59	2.12	1.79–2.52	1.93	1.60–2.32
60+	3.92	3.35–4.60	3.54	2.97–4.21
Dense breasts	0.93	0.81–1.05	1.04	0.90–1.20
Time since last mammogram				
No previous mammogram	1.0	—	1.00	—
1–35 mo	2.51	2.04–3.12	1.97	1.57–2.46
36+ mo	2.11	1.65–2.71	1.87	1.44–2.43
Previous biopsy	1.09	0.96–1.24	0.85	0.74–0.98
Family history of breast cancer	1.37	1.15–1.64	1.20	0.99–1.46
Self-report of breast problem	1.24	0.94–1.64	1.51	1.10–2.09
Ultrasound after screening	0.88	0.77–0.99	0.99	0.87–1.14

Note—Dash (—) indicates the reference group.

^a Adjusted for age, breast density, history of breast cancer in a first-degree relative, history of breast procedure, interval since last mammogram, and use of ultrasound during workup of the abnormal screening mammographic finding.