Published in final edited form as:
Breast J. 2012 ; 18(4): 326–333. doi:10.1111/j.1524-4741.2012.01250.x.

Reproducibility of BI-RADS Breast Density Measures Among Community Radiologists: A Prospective Cohort Study

Mary C. Spayne, MPH[1,*], **Charlotte C. Gard, MS**[2,3], **Joan Skelly, MS**[4], **Diana L. Miglioretti, PhD**[2,3], **Pamela M. Vacek, PhD**[4], and **Berta M. Geller, EdD**[5]

[1]Canadian Partnership Against Cancer, Toronto, ON

[2]Biostatistics Unit, Group Health Research Institute, Group Health Cooperative, Seattle, WA

[3]Department of Biostatistics, University of Washington School of Public Health, Seattle, WA

[4]Medical Biostatistics, University of Vermont, Burlington, VT

[5]Departments of Family Medicine and Radiology, University of Vermont, Burlington, VT

SUMMARY

Using data from the Vermont Breast Cancer Surveillance System, we studied the reproducibility of Breast Imaging Reporting and Data System (BI-RADS) breast density among community radiologists interpreting mammograms in a cohort of 11,755 postmenopausal women. Radiologists interpreting two or more film-screen screening or bilateral diagnostic mammograms for the same woman within a 3–24 month period during 1996–2006 were eligible. We observed moderate to substantial overall intra-rater agreement for use of BI-RADS breast density in clinical practice, with an overall intra-radiologist percent agreement of 77.2% (95% confidence interval (CI), 74.5–79.5%), an overall simple kappa of 0.58 (95% CI, 0.55–0.61), and an overall weighted kappa of 0.70 (95% CI, 0.68–0.73). Agreement exhibited by individual radiologists varied widely, with intra-radiologist percent agreement ranging from 62.1–87.4% and simple kappa ranging from 0.19–0.69 across individual radiologists. Our findings underscore the need for further evaluation of the BI-RADS breast density categorization system in clinical practice.

Keywords

BI-RADS breast density; intra-radiologist agreement; mammography; clinical practice

INTRODUCTION

John Wolfe established an association between mammographic breast density and risk of breast cancer in a 1976 publication [1]. Since then, several studies have established breast density as an important risk factor for developing the disease [2–13], reporting a 1.8 to 6.0 times greater risk of breast cancer in women with dense breast tissue compared to those with lucent breast tissue. Breast density has emerged as an independent risk factor for breast cancer [9,11], with relative risks exceeded or equaled only by age, BRCA gene mutation, and prior history of breast cancer or atypical hyperplasia [14].

Correspondence: Berta M. Geller, Ed.D., Research Professor, Health Promotion Research, University of Vermont, 429AR4, One South Prospect St. Elevator C-4426, Burlington, VT 05401-3444, Phone: (802) 656-4115, Fax: (802) 656-8826, berta.geller@uvm.edu.

*The basis of this study was drawn from a Master's Thesis undertaken at the University of Massachussetts, Amherst, MA

The study design, analysis, and interpretation of the data are the sole responsibility of the authors.

In the United States the most common method of reporting breast density clinically uses the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) [15]: 1, almost entirely fat; 2, scattered fibroglandular densities; 3, heterogeneously dense; and 4, extremely dense. BI-RADS density was originally introduced to let radiologists record their level of concern that dense tissue might obscure a cancer on mammography [16]. However, lack of clear definition of these categories has hindered reproducibility of BI-RADS density measures [17]. In 2003, to improve standardization and reproducibility, the ACR added percent glandular tissue to each category: 1, <25%; 2, 25–50%; 3, 51–75%; and 4, >75% [18]. Several computer-assisted methods for measuring breast density exist but are not in widespread use [19]; some are too time consuming and labor intensive to be practical for screening in clinical practice [17,20].

Reproducibility of breast density has not been widely studied, and the few studies undertaken have assessed different measurement methods, impeding comparison across studies [20–26]. Most studies of reproducibility have included few radiologists, reading few mammograms twice in blinded order. Rates of reproducibility in studies vary by density classification systems and measurement methods.

To our knowledge, no studies have assessed the reproducibility of BI-RADS breast density categories in the clinical setting where community radiologists interpreted multiple screening mammograms on the same women. Recently developed models for predicting breast cancer risk [27–29] have included BI-RADS density. As such models start being used in clinical practice—to guide decisions about how often a woman is screened or use of chemoprevention, for example—it will necessitate understanding how often BI-RADS density measurements are misclassified. This study prospectively followed a cohort of postmenopausal women receiving mammograms in Vermont from 1996 to 2006, with the BI-RADS system in wide use. By examining how consistently individual community radiologists rated breast density, we seek to understand better how well the BI-RADS breast density classification system works in clinical practice.

MATERIALS AND METHODS

Study Population

Data for the study come from the Vermont Breast Cancer Surveillance System (VBCSS) [30], which participates in the National Cancer Institute's Breast Cancer Surveillance Consortium (BCSC) [31]. Routine collection of BI-RADS breast density measures began in 1996. Data were sent to the BCSC's Statistical Coordinating Center for analysis.

Women were eligible for the study if they were postmenopausal, had no history of breast cancer, and had two or more film-screen screening or bilateral diagnostic mammograms including BI-RADS breast density assessments between January 1, 1996 and December 31, 2006. Breast density may appear different on film-screen compared to digital mammography [32]; therefore, we excluded digital mammograms. Women were considered postmenopausal if they were aged 55 years or older or reported having experienced natural menopause, having had both ovaries removed, or having more than 365 days elapse since their last menstrual period. We excluded premenopausal and perimenopausal women and those under age 50 because breast density may vary during phases of the menstrual cycle [33,34], and breast density may decline by as much as 20% during the menopausal transition [9]. By limiting our study to postmenopausal women over age 50, we focused on subjects whose breast density was expected to remain relatively stable over the 3–24 month time period examined. We excluded women with a history of breast cancer, because therapeutic measures undertaken as a result of the disease may alter breast density [35,36]. Lastly, because hormone therapy (HT) may increase breast density [37,38], and tamoxifen may

reduce it [39,40], we excluded mammograms after self-report of HT, tamoxifen, or raloxifene use, as reported in the health history questionnaire at time of mammogram.

To evaluate the intra-rater agreement in assessment of breast density, we identified all radiologists who interpreted two or more film-screen mammograms for the same woman within a 3–24 month period between 1996 and 2006. While the American Cancer Society recommends yearly screening for women aged 50–69 years [41], a large population-based study where physicians also recommended yearly screening found the median time between screenings in this age group to be 17.7 months [42], approaching the two-year interval for that age group now advised by the U.S. Preventive Services Task Force [43]. In choosing a 3–24 month period for study, our goal was to capture the average screening behavior of postmenopausal women, minimizing time between interpretations, thereby reducing the likelihood of a noticeable change in breast density, while attempting to ensure that interpretations were independent. For each woman, we identified all pairs of mammograms that the same radiologist interpreted 3–24 months apart during 1996–2006. If a radiologist interpreted more than one pair of mammograms for a given woman during the study period, we selected the pair closest in date for that radiologist. For women with multiple pairs of mammograms interpreted by different radiologists, we randomly selected one pair for analysis. In addition, to increase the precision of individual kappa estimates, we required that study radiologists have more than 100 records (mammogram pairs).

The Institutional Review Boards at the University of Vermont and the Group Health Research Institute (home of the Statistical Coordinating Center) approved this study, which complied with the Health Insurance Portability and Accountability Act (HIPAA). The VBCSS and the Statistical Coordinating Center have received a Federal Certificate of Confidentiality and other protection for the identities of women, physicians, and facilities who are subjects of this research.

Statistical Analysis

We used percent agreement and simple [44] and weighted [45] kappa statistics to measure intra-rater agreement in BI-RADS assessment of breast density. Weights for the kappa coefficient were computed following the Fleiss-Cohen method [46]. Landis and Koch [47,48] interpret kappa values as follows: values of <0 represent poor agreement; 0.00–0.20 represent slight agreement; 0.21–0.40 represent fair agreement; 0.41–0.60 represent moderate agreement; 0.61–0.80 represent substantial agreement; and 0.81–1.00 represent almost perfect agreement.

We calculated overall agreement based on cross-tabulations of BI-RADS density measures at the first and second assessments. For category-specific agreement, cross-tabulations were based on measurements for a given category versus measurements for all other categories combined. To account for within-radiologist correlation, we used the bootstrap method [49] to construct confidence intervals for estimates of agreement, re-sampling at the radiologist level. We estimated 95% confidence intervals from the 2.5% and 97.5% percentiles of 2,500 bootstrap samples.

Because we might expect breast density to change correspondingly in women whose body mass index (BMI) changed from first to second assessment, we performed sensitivity analyses, restricting to women whose BMI at first and second density assessments were non-missing, stratifying by whether BMI differed by <1 kg/m^2, <2 kg/m^2, or <4 kg/m^2. We also performed sensitivity analyses to assess the impact of the November 2003 BI-RADS density definition change, restricting to women for whom both density assessments were before November 1, 2003 and to those for whom both density assessments were made after May 30, 2004 (allowing six months for the definition change to take effect).

Bootstrapping and plotting were performed using Stata/SE 9.2 (StataCorp, College Station, TX). All remaining analyses were performed using SAS 9.1 (SAS Institute, Cary, NC).

RESULTS

In the study 34 radiologists interpreted between 119 and 1,033 mammogram pairs on a total of 11,755 postmenopausal women. Women were predominantly white (98%) with median age of 66 years (interquartile range (IQR), 58–73 years) and median BMI of 26.6 kg/m^2 (IQR, 23.4–30.7 kg/m^2) (Table 1).

The distributions of BI-RADS density measures at first and second breast density assessments were very similar, with approximately 9.5% of women placed in category 1 or "almost entirely fat," 60.5% in category 2 or "scattered fibroglandular densities," 27.5% in category 3 or "heterogeneously dense," and 2.5% in category 4 or "extremely dense" (Table 2).

Similar total numbers of women were placed in each category at first and second assessments, but the breast density of 22.8% of women was interpreted differently at first and second readings. Of 1,156 women interpreted at first reading as category 1, at second reading only 634 (54.8%) were interpreted as category 1 and 503 (43.5%) as category 2 (Table 2). Of 7,174 women interpreted at first reading as category 2, at second reading 5,932 (82.7%) were interpreted as category 2,430 (6.0%) as category 1, and 789 (11.0%) as category 3. Of 3,126 women interpreted at first reading as category 3, at second reading 2,365 (75.7%) were interpreted as category 3 and 624 (20.0%) as category 2. Of 299 women interpreted at first reading as category 4, at second reading 148 (49.5%) were interpreted as category 4 and 132 (44.2%) as category 3. Percent agreement was higher for categories 1 and 4 (91.8% and 97.4%, respectively) than for categories 2 and 3 (79.7% and 85.6%, respectively). First and second readings differed by more than one category for less than 1% of women.

Overall, between the first and second breast density assessments, percent agreement was 77.2% (95% confidence interval (CI), 74.5–79.5%); simple kappa was 0.58 (95% CI, 0.55–0.61), representing moderate agreement; and weighted kappa was 0.70 (95% CI, 0.68–0.73), representing substantial agreement (Table 3). Percent agreement ranged from 62.1–87.4% across individual radiologists, simple kappa ranged from 0.19–0.69, and weighted kappa ranged from 0.21–0.79. Six of the 34 (18%) radiologists had only slight or fair agreement, 20 (59%) had moderate agreement, and 8 (24%) had substantial agreement. We saw no evidence of an association between agreement and number of pairs of interpretations per radiologist (Figure 1). Estimates of agreement for women whose BMI at first and second density assessments differed by <1 kg/m^2 (N = 6,446 women, 34 radiologists), by <2 kg/m^2 (N = 8,489, 34 radiologists), and by <4 kg/m^2 (N=9,648, 34 radiologists) were comparable to those for overall agreement, as were estimates for women for whom both density assessments were made before November 1, 2003 (N=8,269, 33 radiologists). Estimates of agreement were slightly higher for women for whom both density assessments were made after May 30, 2004 (N=1,623 women, 25 radiologists).

The median time between breast density assessments was 13.1 months (IQR, 12.2–15.5 months). Intra-rater agreement for interpretations of breast density varied little by time between assessments (Table 3). When stratifying by age at first assessment, agreement was substantial for each 5-year age group but decreased as age increased from 50–54 years [weighted kappa = 0.77 (95% CI, 0.73–0.80)] to 70–74 years [weighted kappa = 0.65 (95% CI, 0.60–0.69)]. Percent agreement ranged from 79.5% (95% CI, 76.9–82.0%) for women aged 50–54 years to 75.9% (95% CI, 71.7–79.4%) for women aged 70–74 years.

DISCUSSION

Our study found moderate to substantial overall intra-rater agreement, but wide variability among radiologists in the reproducibility of their breast density measurements, with 18% having only slight or fair agreement. Our study is the first to assess intra-rater agreement in BI-RADS breast density measures in a large screening population where the basis for agreement is two consecutive mammograms for individual women whose breast density is expected to remain relatively stable. Therefore, our study is difficult to compare directly with prior studies, which based reproducibility on repeat interpretations of the same mammogram in a study setting.

Kerlikowske et al. [23] assessed intra-rater agreement of BI-RADS breast density among community radiologists in a large screening population based on repeat interpretations of the same mammogram. The authors reported substantial overall agreement, with simple kappa equal to 0.72 (95% CI, 0.66–0.78). Our study found lower intra-rater agreement than did Kerlikowske et al., which was based on a total of 790 BI-RADS density interpretations by two radiologists who participated in two training sessions before the study, during which BI-RADS terms were reviewed and interpretations were compared and discussed. Radiologists in our study underwent no special training.

Ciatto et al. [26] reported an average intra-observer kappa of 0.71 based on 12 radiologists' interpretations of 100 digitized mammograms from an existing test set. Simple kappa for individual radiologists ranged from 0.32 to 0.88. Compared with Ciatto et al., our study found a lower overall intra-observer agreement and a narrower range for individual radiologists. This may be because in Ciatto et al., second interpretations followed first interpretations by a median of only 10 days, with mammograms presented to radiologists in the same order at the first and second interpretations, and with radiologists given ACR recommendations for reporting breast density before interpreting.

Of interest in our study is the shift of BI-RADS assessments between categories. Breast density tends to decrease in postmenopausal women by 1–2% per year [50], and for roughly a third of women in our study the time between interpretations was between 13–18 months. However, we observed an upward shift from category 1 to category 2 of 44%, and a shift from category 2 to category 3 of 11%. This illustrates that shift in second interpretation in these categories did not result from natural loss of breast density. In addition, almost half of women assessed as category 4 at first interpretation shifted to category 3 at second interpretation, further suggesting that breast density may be more likely interpreted differently between mammograms due to individual radiologists' misclassification and/or inconsistency, rather than biological or visual change. Prior studies [20,51] assessing inter-radiologist agreement of BI-RADS density have evidenced greatest agreement for radiologists assessing the least dense and most dense categories of breast density. However, we found that only about half of women interpreted as both extreme categories of 1 and 4 at first assessment were interpreted as categories 1 and 4 at second assessment.

Of concern is how the radiologist defines each category of breast density. For example, the ACR BI-RADS 4th edition defines category 1 as <25% glandular tissue. If breast density approaches 25%, the same radiologist may interpret it as either category 1 or category 2 at different times, without any actual change in density occurring. Because breast density is assessed ordinally, we cannot know how much of the drift between categories is due to biological change and how much to radiologist misclassification. If breast density were interpreted predominantly as a continuous variable, these distinctions could be made more easily. Only one automated quantitative measurement tool of breast density is currently available for clinical use [54].

Our study has several limitations. Breast density is generally reported because it can influence accuracy, not because it is a breast cancer risk factor. Mammograms of dense breasts have lower sensitivity and specificity than mammograms of fatty breasts [56,57]. How radiologists consider and report breast density may differ because of their primary goal for reporting. Clinical practice dictates that each new mammogram is assessed for changes in comparison to previous mammograms. On second mammogram, every radiologist in this study has had at least one prior mammogram to refer to for individual women; reproducibility may be expected to benefit from this practice, resulting in artificially inflated agreement. Even so, agreement is lower in our study than in others. In addition, we could not explore radiologists' experience or volume of mammograms interpreted, both of which may influence agreement. The study encompasses 11 years, during which technological changes have occurred, possibly affecting the assessment of breast density. In addition, in November 2003, the ACR added percent density to the descriptions of each BI-RADS category, and this may have influenced how radiologists subsequently categorized density. However, we found little difference between intra-rater agreement based on density interpretations before and after the definition change. Lastly, our study assesses only intra-rater agreement and is restricted to radiologists interpreting two or more mammograms for the same woman; in clinical practice different radiologists may interpret a woman's mammograms over time. Estimates of both intra- and inter-radiologist agreement are needed to determine whether clinically-determined BI-RADS breast density measurements are useful for assessing breast cancer risk for individual women.

In conclusion, our study found individual community radiologists varied considerably in the reproducibility of their BI-RADS breast density interpretations. Because breast density is a notable risk factor for developing breast cancer and is beginning to be used in predicting breast cancer risk for individual women, it is important to quantify it as accurately as possible [27–29]. Our study results suggest that categories of BI-RADS breast density should be defined more clearly and adhered to more rigorously if it is to be useful for clinical purposes such as risk prediction.

A parallel step in continuing to improve the reproducibility of BI-RADS breast density is continued monitoring of how community radiologists use the categories and, when interpretation is found to be variable, implementation of educational and/or quality assurance measures. Prior studies, in which radiologists received instruction regarding use of BI-RADS density categories reported higher intra-rater agreement [23,26]. Given the wide variability in agreement we observed among individual radiologists, it may be useful to examine how much intra-radiologist agreement depends on radiologist characteristics. Such information may be useful in identifying radiologists who are candidates for further training regarding BI-RADS breast density assessment.

Acknowledgments

Cooperative agreements U01CA70013 and U01CA86076 with the National Cancer Institute funded this research.

The authors thank Rebecca Hughes for manuscript editing. We thank the participating VBCSS mammography facilities and radiologists for the data they have provided for this study.

References

1. Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer. 1976; 37:2486–2492. [PubMed: 1260729]
2. Boyd NF, O'Sullivan B, Campbell JE, et al. Mammographic signs as risk factors for breast cancer. Br J Cancer. 1982; 45:185–193. [PubMed: 7059469]

3. Brisson J, Merletti F, Sadowsky NL, Twaddle JA, Morrison AS, Cole P. Mammographic features of the breast and breast cancer risk. Am J Epidemiol. 1982; 115:428–437. [PubMed: 7064977]
4. Brisson J, Morrison AS, Kopans DB, et al. Height and weight, mammographic features of breast tissue, and breast cancer risk. Am J Epidemiol. 1984; 119:371–381. [PubMed: 6702813]
5. Wolfe JN, Saftlas AF, Salane M. Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study. AJR Am J Roentgenol. 1987; 148:1087–1092. [PubMed: 3495132]
6. Brisson J, Verreault R, Morrison AS, Tennina S, Meyer F. Diet, mammographic features of breast tissue, and breast cancer risk. Am J Epidemiol. 1989; 130:14–24. [PubMed: 2545096]
7. Maskarinec G, Meng L. A case-control study of mammographic densities in Hawaii. Breast Cancer Res Treat. 2000; 63:153–161. [PubMed: 11097091]
8. Saftlas AF, Hoover RN, Brinton LA, et al. Mammographic densities and risk of breast cancer. Cancer. 1991; 67:2833–2838. [PubMed: 2025849]
9. Boyd NF, Byng JW, Jong RA, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. J Natl Cancer Inst. 1995; 87:670–675. [PubMed: 7752271]
10. Kato I, Beinart C, Bleich A, Su S, Kim M, Toniolo PG. A nested case-control study of mammographic patterns, breast volume, and breast cancer (New York City, NY, United States). Cancer Causes Control. 1995; 6:431–438. [PubMed: 8547541]
11. Byrne C, Schairer C, Wolfe J, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst. 1995; 87:1622–1629. [PubMed: 7563205]
12. van Gils CH, Hendriks JH, Holland R, et al. Changes in mammographic breast density and concomitant changes in breast cancer risk. Eur J Cancer Prev. 1999; 8:509–515. [PubMed: 10643940]
13. Lam PB, Vacek PM, Geller BM, Muss HB. The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. Cancer. 2000; 89:369–375. [PubMed: 10918168]
14. Singletary SE. Rating the risk factors for breast cancer. Ann Surg. 2003; 237:474–482. [PubMed: 12677142]
15. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS). 2. American College of Radiology; Reston, VA: 1995.
16. Yaffe MJ. Mammographic density - measurement of mammographic density. Breast Cancer Res. 2008; 10:209. [PubMed: 18598375]
17. Harvey JA, Bovbjerg VE. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. Radiology. 2004; 230:29–41. [PubMed: 14617762]
18. American College of Radiology. Breast Imaging Reporting and Data System. 4. Reston (VA): American College of Radiology; 2003.
19. Shepherd JA, Herve L, Landau J, Fan B, Kerlikowske K, Cummings SR. Novel use of single X-ray absorptiometry for measuring breast density. Technol Cancer Res Treat. 2005 Apr; 4(2):173–182. [PubMed: 15773786]
20. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. AJR Am J Roentgenol. 2000; 174:1769–1777. [PubMed: 10845521]
21. Benichou J, Byrne C, Capece LA, et al. Secular stability and reliability of measurements of the percentage of dense tissue on mammograms. Cancer Detect Prev. 2003; 27:266–274. [PubMed: 12893074]
22. Prevrhal S, Shepherd JA, Smith-Bindman R, Cummings SR, Kerlikowske K. Accuracy of mammographic breast density analysis: results of formal operator training. Cancer Epidemiol Biomarkers Prev. 2002; 11:1389–1393. [PubMed: 12433716]
23. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. J Natl Cancer Inst. 1998; 90:1801–1809. [PubMed: 9839520]

24. Jong R, Fishell E, Little L, Lockwood G, Boyd NF. Mammographic signs of potential relevance to breast cancer risk: the agreement of radiologists' classification. Eur J Cancer Prev. 1996; 5:281–286. [PubMed: 8894565]
25. Boyd NF, Wolfson C, Moskowitz M, et al. Observer variation in the classification of mammographic parenchymal pattern. J Chronic Dis. 1986; 39:465–472. [PubMed: 3711253]
26. Ciatto S, Houssami N, Apruzzese A, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. The Breast. 2005; 14:269–275. [PubMed: 16085233]
27. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med. 2008; 148:337–347. [PubMed: 18316752]
28. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst. 2006; 98:1204–14. [PubMed: 16954473]
29. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. Breast Cancer Res Tr. 2005; 94:115–122.
30. Geller BM, Worden JK, Ashley JA, Oppenheimer RG, Weaver DL. Multipurpose statewide breast cancer surveillance system: the Vermont experience. J Regist Manage. 1996; 23:168–174.
31. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol. 1997; 169:1001–1008. [PubMed: 9308451]
32. Harvey JA. Quantitative assessment of percent breast density: analog versus digital acquisition. Technol Cancer Res Treat. 2004 Dec; 3(6):611–616. [PubMed: 15560719]
33. White E, Velentgas P, Mandelson MT, et al. Variation in mammographic breast density by time in menstrual cycle among women aged 40–49 years. J Natl Cancer Inst. 1998; 90:906–910. [PubMed: 9637139]
34. Ursin G, Parisky YR, Pike MC, Spicer DV. Mammographic density changes during the menstrual cycle. Cancer Epidemiol Biomarkers Prev. 2001; 10:141–142. [PubMed: 11219771]
35. Dershaw DD, Shank B, Reisinger S. Mammographic findings after breast cancer treatment with local excision and definitive irradiation. Radiology. 1987; 164:455–461. [PubMed: 3037592]
36. Dershaw DD. Evaluation of the breast undergoing lumpectomy and radiation therapy. Radiol Clin North Am. 1995; 33:1147–1160. [PubMed: 7480662]
37. Rutter CM, Mandelson MT, Laya MB, Seger DJ, Taplin S. Changes in breast density associated with initiation, discontinuation, and continuing use of hormone replacement therapy. JAMA. 2001; 285:171–176. [PubMed: 11176809]
38. Lundstrom E, Wilczek B, von Palffy Z, Soderqvist G, von Schoultz B. Mammographic breast density during hormone replacement therapy: effects of continuous combination, unopposed transdermal and low-potency estrogen regimens. Climacteric. 2001; 4:42–48. [PubMed: 11379377]
39. Brisson J, Brisson B, Cote G, Maunsell E, Berube S, Robert J. Tamoxifen and mammographic breast densities. Cancer Epidemiol Biomarkers Prev. 2000; 9:911–915. [PubMed: 11008908]
40. Atkinson C, Warren R, Bingham SA, Day NE. Mammographic patterns as a predictive biomarker of breast cancer risk: effect of tamoxifen. Cancer Epidemiol Biomarkers Prev. 1999; 8:863–866. [PubMed: 10548313]
41. Leitch AM, Dodd GD, Costanza M, et al. American Cancer Society guidelines for the early detection of breast cancer: update 1997. CA Cancer J Clin. 1997; 47:150–153. [PubMed: 9152172]
42. Ulcickas Yood M, McCarthy BD, Lee NC, Jacobsen G, Johnson CC. Patterns and characteristics of repeat mammography among women 50 years and older. Cancer Epidemiol Biomarkers Prev. 1999; 8:595–599. [PubMed: 10428196]
43. US Preventive Services Task F. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009; 151(10):715–726.
44. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20:37–46.

45. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull. 1968; 70:213–220. [PubMed: 19673146]
46. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coeffiecient as measures of reliability. Educ Psychol Meas. 1973; 33:613–619.
47. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–174. [PubMed: 843571]
48. Landis JR, Koch GG. A one-way component of variance model for categorical data. Biometrics. 1977; 33:671–679.
49. Efron, B.; Tibshirani, RJ. An introduction to the bootstrap. 2. New York: Chapman and Hall/CRC; 1998.
50. Kopans, DB. Breast Imaging. 2. Lippincott-Raven; Philadelphia PA: 1998. p. 235
51. Ooms EA, Zonderland HM, Eijkemans MJC, et al. Mammography: Interobserver variability in breast density assessment. The Breast. 2007; 16:568–576. [PubMed: 18035541]
52. Lee-Han H, Cooke G, Boyd NF. Quantitative evaluation of mammographic densities: a comparison of methods of assessment. Eur J Cancer Prev. 1995; 4:285–292. [PubMed: 7549820]
53. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol. 1994; 39:1629–1638. [PubMed: 15551535]
54. [accessed April 23, 2012] http://www.hologic.com/en/breast-screening/volumetric-assessment/
55. Byng JW, Yaffe MJ, Jong RA, et al. Analysis of mammographic density and breast cancer risk from digitized mammograms. Radiographics. 1998; 18:1587–1598. [PubMed: 9821201]
56. Patella A, Marziani R, Schippa A, Benedetti S, Mossa S, Mossa B. Breast density changes associated with hormone replacement therapy in postmenopausal women. Effects on the specificity and sensitivity of mammographic screening. Eur J Gynaecol Oncol. 2005; 26(5):485–490. [PubMed: 16285562]
57. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. Review. J Natl Cancer Inst. 2000 Jul 5; 92(13):1081–1087. [PubMed: 10880551]

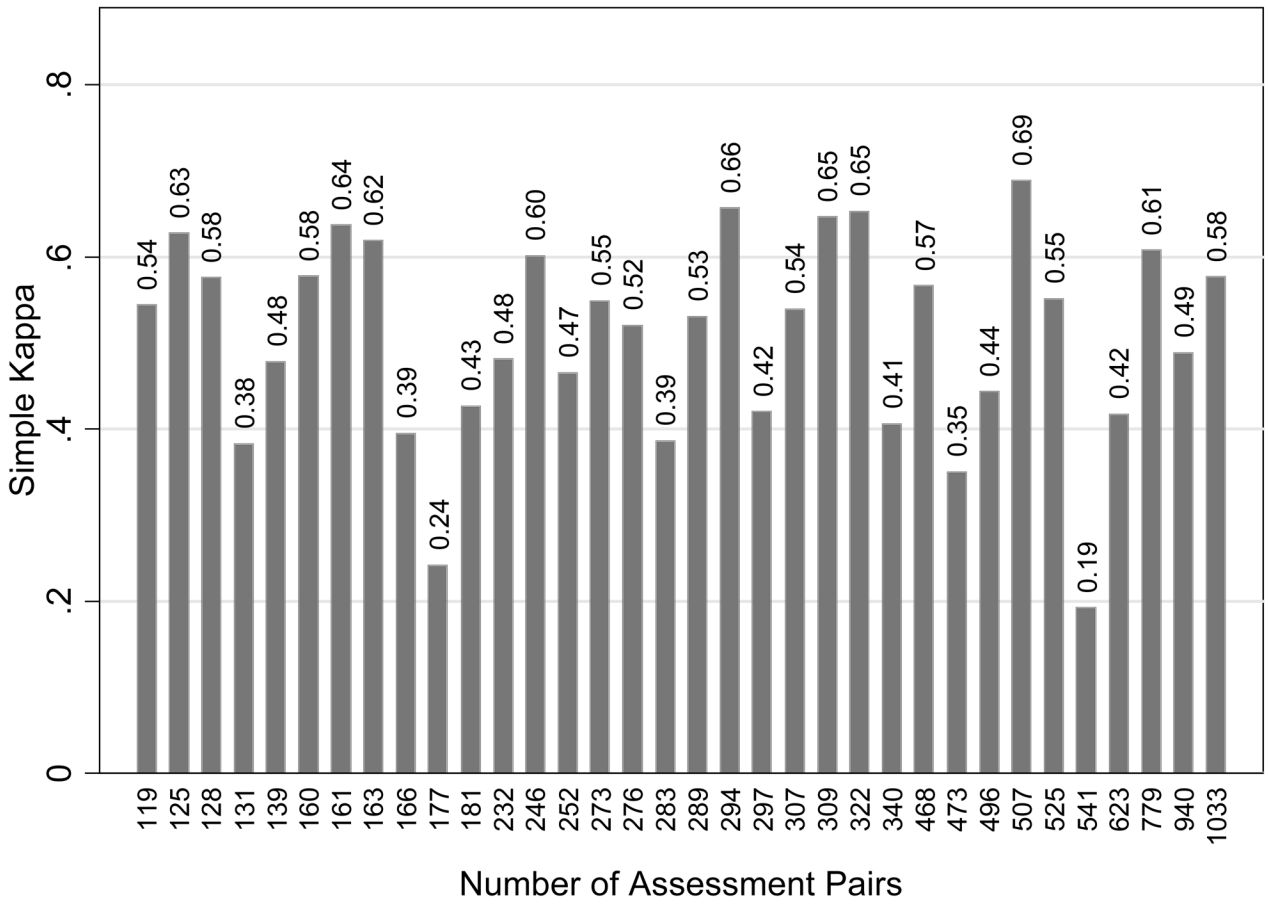

Figure 1.
Simple kappa versus number of assessment pairs per radiologist.

Table 1

Selected characteristics of women at first BI-RADS breast density assessment.

	N	%
	11,755	**100**
Race		
White	11,522	98.0
Non-white	219	1.9
Missing	14	0.1
Age (years)		
50–54	1,364	11.6
55–59	2,221	18.9
60–64	1,884	16.0
65–69	1,942	16.5
70–74	1,889	16.1
≥75	2,455	20.9
Body mass index (kg/m^2)		
<18.5 (underweight)	150	1.3
18.5–24.9 (normal)	3,887	33.1
25–29.9 (overweight)	3,646	31.0
≥30 (obese)	3,009	25.6
Missing	1,063	9.0

Table 2

Cross tabulation of breast density at first and second BI-RADS breast density assessments.

First assessment	Second assessment (*N*, row %)				
	1	**2**	**3**	**4**	**Total**
1	634 (54.8%)	503 (43.5%)	17 (1.5%)	2 (0.2%)	1,156 (9.8%)
2	430 (6.0%)	5,932 (82.7%)	789 (11.0%)	23 (0.3%)	7,174 (61.0%)
3	12 (0.4%)	624 (20.0%)	2,365 (75.7%)	125 (4.0%)	3,126 (26.6%)
4	0 (0.0%)	19 (6.4%)	132 (44.2%)	148 (49.5%)	299 (2.5%)
Total	1,076 (9.2%)	7,078 (60.2%)	3,303 (28.1%)	298 (2.5%)	11,755

Table 3

Intra-rater agreement in assessment of BI-RADS breast density.

	N	Percent agreement (95% CI)	Simple kappa (95% CI)	Weighted kappa (95% CI)
Overall	11,755	77.2 (74.5, 79.5)	0.58 (0.55, 0.61)	0.70 (0.68, 0.73)
Change in body mass index (BMI) <1 kg/m^2	6,446	77.2 (74.2, 79.7)	0.59 (0.55, 0.63)	0.71 (0.68, 0.73)
Change in BMI <2 kg/m^2	8,489	77.3 (74.6, 79.7)	0.59 (0.55, 0.62)	0.71 (0.68, 0.73)
Change in BMI <4 kg/m^2	9,648	77.2 (74.4, 79.8)	0.59 (0.55, 0.62)	0.71 (0.68, 0.73)
Time between assessments				
3 to <12 months	1,005	78.1 (74.0, 81.3)	0.56 (0.51, 0.61)	0.67 (0.60, 0.73)
12–18 months	9,148	77.3 (74.5, 79.4)	0.59 (0.55, 0.62)	0.71 (0.68, 0.73)
19–24 months	1,602	76.5 (73.1, 79.8)	0.57 (0.52, 0.62)	0.70 (0.66, 0.73)
Age (years) at first assessment				
50–54	1,364	79.5 (76.9, 82.0)	0.66 (0.61, 0.69)	0.77 (0.73, 0.80)
55–59	2,221	77.3 (74.7, 79.7)	0.60 (0.56, 0.64)	0.72 (0.68, 0.76)
60–64	1,884	77.5 (74.2, 80.4)	0.58 (0.53, 0.62)	0.70 (0.66, 0.74)
65–69	1,942	76.0 (73.2, 78.7)	0.55 (0.50, 0.58)	0.66 (0.61, 0.70)
70–74	1,889	75.9 (71.7, 79.4)	0.54 (0.48, 0.59)	0.65 (0.60, 0.69)
≥75	2,455	77.8 (74.3, 81.0)	0.58 (0.53, 0.63)	0.70 (0.66, 0.73)
Relative to change in BI-RADS density definition				
Before November 1, 2003	8,269	77.3 (74.7, 79.5)	0.58 (0.53, 0.61)	0.70 (0.66, 0.72)
After May 30, 2004	1,623	79.1 (75.0, 82.4)	0.64 (0.59, 0.67)	0.74 (0.67, 0.79)

CI = confidence interval

Breast J. Author manuscript; available in PMC 2013 May 21.