



Published in final edited form as:

Med Care. 2011 January ; 49(1): 67–75. doi:10.1097/MLR.0b013e3181f380e0.

Accuracy of Diagnostic Mammography at Facilities Serving Vulnerable Women

L. Elizabeth Goldman, MD, MCR¹, Rod Walker, MS², Diana L. Miglioretti, PhD^{2,3}, Rebecca Smith-Bindman, MD⁴, and Karla Kerlikowske, MD¹ for the National Cancer Institute Breast Cancer Surveillance Consortium

¹ Department of Medicine, University of California San Francisco, San Francisco, CA

² Biostatistics Unit, Group Health Research Institute, Seattle, WA

³ Department of Biostatistics, University of Washington, Seattle, WA

⁴ Department of Radiology, Epidemiology and Biostatistics University of California San Francisco, San Francisco, CA

Abstract

Background—Breast cancer missed on diagnostic mammography may contribute to delayed diagnoses, while false-positive results may lead to unnecessary invasive procedures. Whether accuracy of diagnostic mammography at facilities serving vulnerable women differs from other facilities is unknown.

Objective—To compare the interpretive performance of diagnostic mammography at facilities serving vulnerable women to those serving non-vulnerable women.

Design—We examined 168,251 diagnostic mammograms performed at BCSC facilities from 1999–2005. We used hierarchical logistic regression to compare sensitivity, false positive rates, and cancer detection rates.

Subjects—Women ages 40–80 years undergoing diagnostic mammography to evaluate an abnormal screening mammogram or breast problem.

Measures—Facilities were assigned vulnerability indices according to the populations served based on the proportion of mammograms performed on women with lower educational attainment, racial/ethnic minority status, limited household income, or rural residences.

Results—Sensitivity of diagnostic mammography did not vary significantly across vulnerability indices adjusted for patient-level characteristics, but false-positive rates for diagnostic mammography examinations to evaluate a breast problem were higher at facilities serving vulnerable women defined as those with: lower educational attainment (odds ratio (OR) 1.39; 95% confidence interval (CI) 1.08, 1.79); racial/ethnic minorities (OR 1.32; 95% CI 0.98, 1.76); limited income (OR 1.34; 95% CI 1.08, 1.66), and rural residence (OR 1.55; 95% CI 1.27, 1.88).

Conclusions—Diagnostic mammography to evaluate a breast problem at facilities serving vulnerable women has higher false positive rates than at facilities serving non-vulnerable women. This may reflect concerns that vulnerable populations may be less likely to follow-up after abnormal diagnostic mammography or concerns that such populations have higher cancer prevalence.

Keywords

breast cancer; diagnostic mammography; disparities; quality of care

Background

Diagnostic mammography is the principal imaging tool used to diagnosis breast cancer. Accuracy in diagnostic mammography interpretations varies nationally across facilities,(1) in part due to differences in radiologists' experience, equipment, practice patterns, and patient populations undergoing diagnostic mammography. Previous research has focused primarily on the woman and radiologist, factors that contribute to accuracy in diagnostic mammography.(2–5) Disparities in cancer care and outcomes for women with lower education attainment, racial or ethnic minorities, limited income, and those who live in rural areas are well established.(6–8) Whether the interpretative performance of diagnostic mammography at facilities serving a large proportion of women with these demographics is similar to performance at other facilities is less well understood.

Vulnerable women such as those with lower educational attainment, racial or ethnic minorities, limited income, or living in a rural area are at higher risk for poor breast cancer outcomes (6–8). Previous research indicates that screening mammography performed at facilities serving high proportions of vulnerable women had higher specificity and similar sensitivity compared to screening at facilities serving non-vulnerable women.(9) In other words, women seen at facilities serving greater numbers of vulnerable women are more likely to have a normal screening mammography examination when they do not have cancer (i.e. higher specificity rates) than at facilities serving vulnerable women; however, there is no difference in the proportion of cancers that are detected. Findings for screening mammography, however, may differ from those in diagnostic mammography, as these types of mammography may require different interpretive skills.(10) While the goal of both screening and diagnostic mammography is to detect breast cancer, the two types of exams involve different patient populations, very different cancer incidence, different numbers of images taken, possible use of different machines and/or technologists, different interpretation protocols (e.g., batch vs. online reading), and different management recommendations for abnormal assessments.

Since it is the accuracy of diagnostic mammography that confers whether a woman will undergo an invasive biopsy necessary to make the diagnosis of cancer, it is important to determine any differences in diagnostic performance between facilities serving different populations. Moreover, it is important for diagnostic mammography to have high sensitivity to avoid diagnostic delay. In this study, we compare the accuracy of diagnostic mammography interpretations at facilities serving vulnerable women to the accuracy at facilities serving non-vulnerable women in the Breast Cancer Surveillance Consortium (BCSC) (1) to understand whether differences in accuracy could partially account for the disparities in breast cancer severity and mortality in vulnerable populations.

Methods

Data Source

Data were pooled from mammography registries across seven states participating in the National Cancer Institute-funded BCSC and include mammography interpretations linked at the patient-level to pathology and tumor registry data. This consortium was formed to evaluate the quality of mammography nationally. The BCSC population has been shown to

be representative of the U.S. population of women with characteristics that are similar to national demographics in terms of age, ethnicity, and urban or rural residence. (1)

The mammography registries prospectively collect women's self-reported demographic information and breast cancer risk factor data at each mammography examination, together with radiologists' reports on screening and diagnostic mammography. The BCSC links to 2000 U.S. Census Bureau data based on the women's zip codes to obtain population-level socio-demographic information. Registries ascertain cancer outcomes through linkage with state tumor registries or regional Surveillance, Epidemiology, and End Results (SEER) programs, as well as linkages to pathology databases at five of the seven mammography registries. Each registry and the Statistical Coordinating Center have received IRB approval for either active or passive consenting processes or a waiver of consent to enroll participants, link data, and perform analytic studies. All procedures are Health Insurance Portability and Accountability Act (HIPAA) compliant and all registries 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.

Study subjects

We included women ages 40–80 years who underwent at least one diagnostic mammography examination between January 1999 and December 2005 identified by the radiologist as either performed for the evaluation of a recent abnormal screening result (termed 'additional evaluation of a recent abnormal screening result') or performed for the 'evaluation of a symptomatic breast problem.' Of the 384,063 diagnostic mammography interpretations initially identified from women age 40–80 during the years 1999–2005, we excluded mammography from women with a history of breast cancer [$N=57,786$, 15%] based on self-report or linkage with the cancer registry or pathology databases; those from women who reported presence of breast implants at the time of examination ($N=6,155$, 1.6%), (11), those missing a final mammographic assessment [$N=1,504$, 3.9%], and those with unknown time since last mammography [$N=34,575$, 9.0%]. In addition, some BCSC facilities do not collect Breast Imaging Reporting and Data System (BI-RADS) breast density as it is not required to collect breast density by the Mammography Quality Standards Act or for American College of Radiation accreditation, yet as it is a potential confounding variable related to both interpretive performance and patient characteristics, (5, 12) we excluded mammography interpretations where breast density was not reported [$N=113,280$, 29%].

Diagnostic mammography accuracy

We used standard definitions for diagnostic mammography interpretations based on the final assessment at the end of imaging work-up, which could be up to 90 days following the initial diagnostic examination. (2, 13) BI-RADS is the standard lexicon used for interpreting mammography. (13) Final assessments of BI-RADS 4 (suspicious abnormality) or 5 (highly suggestive of malignancy), or 0 (incomplete) or 3 (probably benign finding) with a recommendation for biopsy, fine needle aspiration (FNA), or surgical consult were classified as positive interpretations. A negative diagnostic mammography examination was defined as a mammography interpretation with a BI-RADS 1 (negative) or 2 (benign finding) or a BI-RADS 0 or 3 without a final recommendation for biopsy/FNA/surgical consult. If the final assessment within 90 days in the BCSC database had a BI-RADS 0 with recommendation for additional imaging, non-specified workup, or a missing recommendation, we considered the assessment to be missing and excluded the mammography from the analysis (0.4%). We considered women to have a diagnosis of breast cancer if the state tumor registries or regional Surveillance, Epidemiology, and End Results (SEER) programs, or the pathology databases showed any invasive carcinoma or ductal carcinoma *in situ* within 12 months of the diagnostic exam. We also included cancers diagnosed within 30 days prior to diagnostic

mammography because some cancer registries define the date of diagnosis to be the first evidence of breast cancer rather than the date of the biopsy confirmation, and this evidence could come from a prior recent abnormal screening result or clinical exam which led to the additional diagnostic evaluation.(1) Sarcomas, lymphomas, and lobular carcinoma *in situ* were not considered breast cancer.

To evaluate accuracy of diagnostic mammography, we calculated sensitivity, false positive rates, and cancer detection rates. (1, 14) Sensitivity was calculated from the number of true-positive mammography examinations (positive final assessments with breast cancer) divided by the number of breast cancers, and false positive rates were calculated from the number of false-positive examinations (positive final assessments without breast cancer) divided by the number of non-breast cancers. Cancer detection rate was defined as the number of cancers detected (true positives) per 1,000 diagnostic mammography interpretations.

Definition of facilities serving vulnerable women

For this analysis, we used the methodology and definitions employed in a prior study that analyzed the accuracy of screening mammography at facilities serving vulnerable populations.(9) Vulnerability was based on four socio-demographic characteristics: educational attainment, race/ethnicity, household income, and whether living in a rural or urban area. We first measured these characteristics for all women. Self-reported information provided at the time of mammography was used to determine a woman's educational attainment and race/ethnicity. For income and rural/urban status, geocoded linkages between 2000 Census data and self-reported residential zip code at the time of mammography were used to assign each woman an income measure corresponding to the median household income in the zip code and a rural/urban score corresponding to the percentage of rural residences in the zip code.

To describe the vulnerability of the population served by each mammography facility, we calculated a continuous facility-level vulnerability index by aggregating individual woman-level characteristics for the four vulnerability measures across all mammography examinations (both screening and diagnostic) served by a given facility during the 1999–2005 study period. The continuous index measures were (1) the percentage of the population with a high school education or less, (2) the percentage of the population composed of minorities (self-reported African-American race, or Hispanic/Pacific-Islander/Hawaiian/ Native American ethnicity), (3) the average median household income and (4) the average percentage of rural residents. We did not include Non-Pacific Islander Asian Americans as a vulnerable minority because their breast cancer mortality rates are lower than that of Caucasians and other minority groups.(15) The continuous index measures of vulnerability were then dichotomized to provide a binary facility-level vulnerable/not vulnerable classification for the population served by each facility. The cutoffs for these dichotomized variables were determined by taking one standard deviation from the means of the continuous vulnerability measures for our study population. Specifically, we classified facilities as serving a vulnerable population if: 1) 17% of mammography interpretations were from women who had not completed high school (lower educational attainment); 2) the percentage minority was > 30% (racial/ethnic minority); 3) the average median income was < \$45,000 (limited income); or 4) the average percentage of rural residences was > 52% (rural residence). We also created a composite facility-level vulnerability score by adding 1 for each of the binary vulnerability indices met; each component was given equal weighting, so the score ranged from 0 to 4. For descriptive purposes in this paper (Table 1), we referred to a 0 composite measure of vulnerability as 'non-vulnerable', a 1 or 2 as 'moderately vulnerable', and a 3 or 4 as 'highly vulnerable', however for the main analysis we included all 5 measures. Finally, each facility's four binary vulnerability indices and its composite index were assigned to every diagnostic mammography examination performed on women

within the facility. Thus, the vulnerability score is a characteristic of the population to which the woman (mammography) belongs, rather than a characteristic of the woman herself.

To ensure interpretability and stability of the facility vulnerability categorizations, we excluded facilities if any of their 4 vulnerability classifications (serving limited education, racial/ethnic minorities, women with limited income, and rural residents) were missing (N = 2) or changed more than twice during the 7 year study period (N = 2).

Analyses

For all analyses, we analyzed separately two types of diagnostic mammography examinations (those indicated as follow-up of a recent abnormal screening result and those indicated for evaluation of a breast problem) because interpretive performance measures of different indications for diagnostic examinations are known to vary significantly. (1) We described the total number of diagnostic mammography interpretations by age group, BI-RADS breast density, time since last mammography, and BI-RADS final assessment, and stratified results by facility vulnerability categories. We then calculated the unadjusted sensitivity, false positive rate and cancer detection rate of diagnostic mammography, stratified by each of the four binary vulnerability indices and the composite index.

Adjusted associations (odds ratios and 95% confidence intervals) between diagnostic accuracy and the facility-level vulnerability measures were estimated using logistic-normal mixed-effects models, (16) with adjusted sensitivity, false positive rate, and cancer detection rates estimated from these models using marginal standardization (also known as predictive margins).(17, 18) Each model was specified at the level of the diagnostic mammography examination, with a facility-level random effect introduced to account for clustering of examinations within facilities, and with additional mammography-level covariates included to adjust for other factors that may influence mammography performance including registry site, a woman's age, time since last mammography, and BI-RADS breast density. Woman-level random effects (to account for multiple mammography interpretations from the same woman) were not considered because of computational constraints due to the large number of woman-level clusters and the potential for risk factor covariates to change across a woman's mammography interpretations; prior experience shows that this adjustment does not change the inferences due to the large number of women and small number of mammography interpretations per women. (19) We performed all analyses at the level of the diagnostic mammography examination. This allowed for interpretation of associations in terms of the impact that the vulnerability of the population served has on mammography accuracy at the level of the examination. Statistical modeling was done using PROC NLMIXED in SAS 9.1 (SAS Institute, Carey, NC).

In addition to the main analyses, we conducted a sensitivity analysis where we ascertained cancers for 2 years (730 days) following the index diagnostic mammography, instead of 1 year (365 days), to address the concern that cancer diagnoses in vulnerable women might be delayed due to less access to care or delayed follow-up. (20, 21) To understand the potential implications on our results of excluding mammograms without breast density reported, we compared whether facilities that reported breast density versus those that did not had different interpretive performance. As a post-hoc analysis, we also calculated unadjusted cancer rates, defined as all cancers ascertained in the cancer registries and pathology databases whether or not they were detected on diagnostic mammogram within 12 months of diagnostic mammography examinations, to measure cancer prevalence at facilities serving vulnerable women compared to non-vulnerable women.

Results

Our study sample included 168,251 diagnostic mammography examinations, including 83,464 examinations performed at 153 facilities to evaluate a recent abnormal screening result among 76,199 women, and 84,787 examinations performed at 176 facilities to evaluate a breast problem among 74,785 women (Table 1). About 60% of all mammography interpretations (both indications) occurred at facilities serving a non-vulnerable population; around 30% occurred at facilities classified as serving a moderately vulnerable population; and 10% occurred at facilities classified as facilities serving a highly vulnerable population. This overall distribution was similar to the distributions seen within the two strata defined by diagnostic indications. Facilities serving more vulnerable populations were more likely to serve older women, women with less dense breasts, and women who were less frequently screened, and facilities serving vulnerable women were more likely to recommend a biopsy and less likely to recommend short-interval follow-up for BI-RADS assessments of 0 or 3 ('needs additional imaging or probably a benign finding').

In unadjusted analyses for both diagnostic mammography indications, facilities serving vulnerable women tended to have lower sensitivity and higher false positive rates across most measures of vulnerability (Table 2). Associations between vulnerability and cancer detection rates tended to differ, however, between the two mammography indications. For diagnostic mammography indicated as an additional evaluation of an abnormal screening result, facilities that served predominately rural residents and racial/ethnic minorities had lower cancer detection rates. However, for diagnostic mammography indicated to evaluate a breast problem, facilities serving vulnerable populations (racial/ethnic minorities, limited income, and rural residence) tended to have higher cancer detection rates (Table 2).

For diagnostic mammography indicated as an evaluation of an abnormal screening, most of the observed differences in raw performance measures were no longer significant after adjusting for time since previous mammography, BI-RADS breast density, registry site, and age (Table 3a). After adjustment, only facilities serving limited income women had significantly higher false positive rates relative to those serving non-vulnerable income populations (OR 1.39; 95% CI 1.13, 1.70). Adjusted sensitivity and cancer detection rates were not significantly different, though they did tend to be lower (non-significant) at facilities serving more vulnerable populations. In contrast, for diagnostic mammography taken to evaluate a breast problem, the differences in false positive rates remained after adjustment (Table 3b). Adjusted false positive rates were higher in 3 of the 4 vulnerability categories: lower educational attainment (OR 1.39; 95% CI 1.08, 1.79); limited income (OR 1.34; 95% CI 1.08, 1.66), and rural residence (OR 1.55; 95% CI 1.27, 1.88, and bordered on statistical significance for the fourth category (race/ethnicity OR 1.32, 95% CI 0.98, 1.76). Furthermore, there was a dose response relationship, with increasing vulnerability composite score associated with higher false positive rates (trend: $p < 0.01$). The cancer detection rate remained significantly higher after adjustment at facilities serving predominately rural populations (OR 1.26; 95% CI 1.05, 1.50), and was borderline significant for race/ethnicity and income (Table 3b). There were no significant differences in the adjusted sensitivity of diagnostic mammography interpretations between facilities.

As a sensitivity analysis, we calculated all measures of diagnostic mammography accuracy based on cancers linked using a two year follow-up window from the time of mammography, and we found no substantial differences in our conclusions for sensitivity, false positive rates, or cancer detection rates compared with using one year. (Data not presented) We did not find any substantial differences in sensitivity or false positive rates between facilities that did or did not report breast density. The only difference between facilities excluded for not reporting breast density was that the cancer detection rates among

mammograms taken for evaluation of a breast problem were lower for excluded facilities that did not report breast density (37.6 per 1,000 vs. 21.6 per 1000 exams). We also calculated cancer rates as a measure of cancer prevalence at facilities serving vulnerable women compared to facilities serving non-vulnerable women and found that the unadjusted cancer rates were higher in facilities serving vulnerable women for diagnostic mammography to evaluate breast symptoms; lower educational attainment (cancer rate per 1,000 exams 43.8 vs. 44.7); racial/ethnic minorities (cancer rate 43.7 vs. 45.6); limited income (cancer rate 41.6 vs. 51.3); rural residence (cancer rate 42.8 vs. 48.1), with increasing cancer rates associated with higher composite scores. (Data not shown)

Discussion

For diagnostic mammography performed to evaluate an abnormal screening result, facilities serving vulnerable women had similar interpretive performance to facilities serving non-vulnerable women. Only facilities serving limited income women, one of the four categories of vulnerable women, had significantly higher false positive rates relative to those serving non-vulnerable income populations. In contrast, for the interpretation of diagnostic mammography to evaluate a symptomatic breast problem, facilities serving a greater proportion of vulnerable women were more likely to recommend a biopsy or surgical consultation among women not subsequently diagnosed with breast cancer compared to facilities that did not serve vulnerable women. We did not find associated differences between facilities in the sensitivity or cancer detection rates. The lack of difference in the sensitivity of diagnostic mammography to evaluate an abnormal screening result or breast problem for vulnerable and non-vulnerable populations is reassuring and indicates that the characteristics of the facilities where women go does not appear to influence cancer detection among those with cancer. However, the higher false positive rates at facilities serving vulnerable women suggest these women may be more likely to receive breast biopsies when they don't have cancer.

Our findings for diagnostic mammography differ from our prior study for screening mammography. (9) The previous study, which also used BCSC data (1998–2004), found radiologists at facilities serving women with lower educational attainment, racial/ethnic minorities, limited income, and rural residences tended to have lower false positive rates (higher specificity). These contradictory results suggest factors driving mammography interpretive performance could differ between screening and diagnostic mammography. For one, differences in cancer prevalence among women undergoing screening and diagnostic mammography may influence radiologists' perception of cancer risk and therefore the likelihood that they would recommend women return for further testing. In settings where availability of diagnostic imaging is limited and where cancer prevalence is low (i.e. low-risk screening population), radiologists may be less likely to recall women for diagnostic mammography. In contrast, in settings where the cancer prevalence is higher, as occurs with diagnostic mammography, radiologists may be concerned that women may not return for follow up evaluation, and therefore may be more likely to recommend a biopsy as opposed to short-interval follow-up, additional diagnostic imaging, or clinical follow-up. To clarify, while radiologists are unlikely to know the likelihood of a given woman's follow-up or cancer risk, their practice patterns are likely to be influenced by the overall follow-up rates and cancer prevalence of the population of women evaluated at the mammography facility. Follow-up rates after screening mammography for women with lower educational attainment, racial/ethnic minorities, and women with limited income are lower than for other women.(22) Similar concerns may exist for diagnostic mammography. In addition, because the unadjusted cancer rates (representative of cancer prevalence) at facilities serving vulnerable women are higher, radiologists may have a greater concern that these women are more likely to have cancer and therefore may recommend more biopsies or surgical follow-

up in symptomatic woman attending these facilities. This increased likelihood to recommend biopsy or surgical follow-up could explain the higher false positive rates at these facilities.

Availability of screening and diagnostic mammography may differ across facilities. Radiologists at these facilities may have different experience in interpreting diagnostic and screening mammography. Facilities serving vulnerable populations may tend proportionately to perform more screening than diagnostic mammography, and therefore, have lower false-positive rates for screening mammography and higher false-positive rates for diagnostic mammography. (23) Facilities serving limited income women were the only type of facility serving vulnerable women that demonstrated higher false positive rates for diagnostic mammography to evaluate an abnormal screening result. These facilities may have specific resource limitations, such as lack of breast ultrasound. (24, 25) We did not have these data available to us for this analysis.

There are several important strengths and limitations to the study. We used a diverse cohort of many facilities across seven sites in the United States representative of community practice and evaluated the impact that the vulnerability of a population which a facility serves has on the accuracy of diagnostic interpretations using multiple characterizations of vulnerable women. While the higher false positive rates seen at facilities serving vulnerable women may lead to more biopsies in women who do not ultimately receive a cancer diagnosis, our study could not determine this specifically because detailed utilization data to determine whether referral for biopsy truly equated to receiving a biopsy were not available for all facilities. Lower biopsy rates could delay cancer diagnoses, and in effect, artificially raise the false positive rates in vulnerable women. It is possible that vulnerable women may be less likely to receive follow up, which could artificially increase false positive rates at facilities serving vulnerable women if some of these women lost-to-follow-up truly had cancer. To address this concern, we conducted a sensitivity analysis that extended the follow-up time for diagnosis from 1 to 2 years and did not find any difference in our findings.

We selected several measures of vulnerability to help identify facilities serving vulnerable women; however other definitions could be considered with different thresholds. Our study was limited to mammograms with BI-RADS breast density reported because breast density is a known confounder of interpretive performance. We found no substantial differences in sensitivity or false positive rates between included facilities and those excluded due to missing breast density values; however, the cancer detection rate was somewhat lower among excluded facilities. We note, though, that a number of these excluded facilities were from large urban centers which, consistent with our main analysis results, would be estimated to have lower cancer detection rates than facilities serving more rural populations.

Finally, this analysis evaluated the impact of differences in mammography facility-level characteristics on diagnostic performance at the level of a woman's mammography exam. We did not control for radiologists' experience, equipment, or practice patterns, which can contribute to diagnostic mammography interpretive performance.(2) Our analysis, however, has significance from the perspective of a woman choosing to undergo a mammography at a given facility. While women may be able to select where her mammography is performed, she did not have the ability to select who will interpret her mammogram at a particular facility. The experience of the collective group of radiologists and the equipment or the practice patterns at a facility were unmodifiable facility characteristics from the perspective of the woman; therefore, we did not adjust for them.

In conclusion, for diagnostic mammography indicated for evaluation of a symptomatic breast problem, facilities serving vulnerable populations, in general, had higher rates of biopsy or surgical consultation recommendations in women who did not have a subsequent diagnosis of cancer than did at facilities serving fewer vulnerable women; however, significant differences in sensitivity were not observed between such facilities. Facilities serving limited income women undergoing diagnostic mammography to evaluate an abnormal screening result additionally demonstrated greater rates of biopsy and surgical consultation referral among women who did not have a subsequent diagnosis of cancer than did facilities serving non-limited income women. Research should be conducted to determine the appropriate thresholds for referring women to biopsy in different clinical situations for optimal cancer yield per biopsy. As accuracy may differ between screening and diagnostic mammography, both should be assessed when evaluating the quality of mammography at facilities. Future research should consider evaluating facility characteristics such as availability of ultrasound and other diagnostic resources to better understand potential modifiers of diagnostic accuracy.

Acknowledgments

This work was supported by the Agency for Health Care Research and Quality, Grant #1 K08 HS018090-01, California Breast Cancer Research Project, Grant #141B-0062, and the National Cancer Institute Breast Cancer Surveillance Consortium (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040). 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 authors take full responsibility in the design of the study, the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript. 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. Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. *Radiology*. 2005; 235:775–790. [PubMed: 15914475]
2. Miglioretti DL, Smith-Bindman R, Abraham L, et al. Radiologist characteristics associated with interpretive performance of diagnostic mammography. *J Natl Cancer Inst*. 2007; 99:1854–1863. [PubMed: 18073379]
3. Jackson SL, Taplin SH, Sickles EA, et al. Variability of interpretive accuracy among diagnostic mammography facilities. *J Natl Cancer Inst*. 2009; 101:814–827. [PubMed: 19470953]
4. Jensen A, Vejborg I, Severinsen N, et al. Performance of clinical mammography: a nationwide study from Denmark. *Int J Cancer*. 2006; 119:183–191. [PubMed: 16450388]
5. Barlow WE, Lehman CD, Zheng Y, et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. *J Natl Cancer Inst*. 2002; 94:1151–1159. [PubMed: 12165640]
6. Chlebowski RT, Chen Z, Anderson GL, et al. Ethnicity and breast cancer: factors influencing differences in incidence and outcome. *J Natl Cancer Inst*. 2005; 97:439–448. [PubMed: 15770008]
7. Dalton SO, During M, Ross L, et al. The relation between socioeconomic and demographic factors and tumour stage in women diagnosed with breast cancer in Denmark, 1983–1999. *Br J Cancer*. 2006; 95:653–659. [PubMed: 16909141]
8. Ayanian JZ, Kohler BA, Abe T, et al. The relation between health insurance coverage and clinical outcomes among women with breast cancer. *N Engl J Med*. 1993; 329:326–331. [PubMed: 8321261]
9. Goldman LE, Haneuse SJ, Miglioretti DL, et al. An Assessment of the Quality of Mammography Care at Facilities Treating Medically Vulnerable Populations. *Med Care*. 2008; 46:701–708. [PubMed: 18580389]

10. Beam CA, Conant EF, Sickles EA. Correlation of radiologist rank as a measure of skill in screening and diagnostic interpretation of mammograms. *Radiology*. 2006; 238:446–453. [PubMed: 16436811]
11. Miglioretti DL, Rutter CM, Geller BM, et al. Effect of breast augmentation on the accuracy of mammography and cancer characteristics. *Jama*. 2004; 291:442–450. [PubMed: 14747501]
12. El-Bastawissi AY, White E, Mandelson MT, et al. Variation in mammographic breast density by race. *Ann Epidemiol*. 2001; 11:257–263. [PubMed: 11306344]
13. BI-RADSTM. Radiology ACo. American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADSTM). 3. Reston; Virginia: 1998.
14. Radiology ACo. ACR Practice Guideline for the Performance of Screening and Diagnostic Mammogram. Resolution. 2008; 24:1–10.
15. Kerlikowske K, Creasman J, Leung JW, et al. Differences in screening mammography outcomes among White, Chinese, and Filipino women. *Arch Intern Med*. 2005; 165:1862–1868. [PubMed: 16157830]
16. Diggle, PHP.; Liang, K.; Zeger, S. Analysis of longitudinal data. Oxford: Oxford University Press; 2002.
17. Lane PW, Nelder JA. Analysis of covariance and standardization as instances of prediction. *Biometrics*. 1982; 38:613–621. [PubMed: 7171691]
18. Graubard BI, Korn EL. Predictive margins with survey data. *Biometrics*. 1999; 55:652–659. [PubMed: 11318229]
19. Miglioretti DL, Heagerty PJ. Marginal modeling of multilevel binary data with time-varying covariates. *Biostatistics*. 2004; 5:381–398. [PubMed: 15208201]
20. Allen JD, Shelton RC, Harden E, et al. Follow-up of abnormal screening mammograms among low-income ethnically diverse women: findings from a qualitative study. *Patient Educ Couns*. 2008; 72:283–292. [PubMed: 18490127]
21. Wujcik D, Fair AM. Barriers to diagnostic resolution after abnormal mammography: a review of the literature. *Cancer Nurs*. 2008; 31:E16–30. [PubMed: 18772653]
22. Strzelczyk JJ, Dignan MB. Disparities in adherence to recommended followup on screening mammography: interaction of sociodemographic factors. *Ethn Dis*. 2002; 12:77–86. [PubMed: 11913611]
23. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. *J Natl Cancer Inst*. 2005; 97:358–367. [PubMed: 15741572]
24. Schaefer FK, Waldmann A, Katalinic A, et al. Influence of additional breast ultrasound on cancer detection in a cohort study for quality assurance in breast diagnosis-analysis of 102,577 diagnostic procedures. *Eur Radiol*. 2009
25. Devolli-Disha E, Manxhuka-Kerliu S, Ymeri H, et al. Comparative accuracy of mammography and ultrasound in women with breast symptoms according to age and breast density. *Bosn J Basic Med Sci*. 2009; 9:131–136. [PubMed: 19485945]

Table 1
 Diagnostic mammography-level characteristics, stratified by radiologist indication and composite vulnerability score

| | Additional evaluation of recent abnormal screening result | | | | | | | | | | | |
|---|---|------|--|--------------------------------|------------------------------------|------|--------------------------------|------|--|--------------------------------|------------------------------------|------|
| | Evaluation of breast problem | | | | | | Evaluation of breast problem | | | | | |
| | Composite vulnerability score* | | | Composite vulnerability score* | | | Composite vulnerability score* | | | Composite vulnerability score* | | |
| | Non-vulnerable (score = 0) | | Moderately vulnerable (score = 1 or 2) | | Highly vulnerable (score = 3 or 4) | | Non-vulnerable (score = 0) | | Moderately vulnerable (score = 1 or 2) | | Highly vulnerable (score = 3 or 4) | |
| | N | % | N | % | N | % | N | % | N | % | N | % |
| <i>Total number of unique women served</i> | 43,562 | | 26,057 | | 6,997 | | 44,939 | | 21,331 | | 8,515 | |
| <i>Total number of diagnostic mammography interpretations</i> | 47,609 | | 28,372 | | 7,483 | | 52,050 | | 23,193 | | 9,544 | |
| <i>Age group, years</i> | | | | | | | | | | | | |
| 40 – 49 | 17,720 | 37.2 | 9,901 | 34.9 | 2,207 | 29.5 | 23,136 | 44.4 | 10,086 | 43.5 | 3,781 | 39.6 |
| 50 – 59 | 15,762 | 33.1 | 9,163 | 32.3 | 2,461 | 32.9 | 15,732 | 30.2 | 6,931 | 29.9 | 2,760 | 28.9 |
| 60 – 69 | 8,393 | 17.6 | 5,599 | 19.7 | 1,743 | 23.3 | 7,921 | 15.2 | 3,811 | 16.4 | 1,850 | 19.4 |
| 70 – 80 | 5,734 | 12.0 | 3,709 | 13.1 | 1,072 | 14.3 | 5,261 | 10.1 | 2,565 | 10.2 | 1,153 | 12.1 |
| <i>Breast density</i> | | | | | | | | | | | | |
| 1: Almost entirely fat | 1,416 | 3.0 | 1,298 | 4.6 | 276 | 3.7 | 3,225 | 6.2 | 1,509 | 6.5 | 493 | 5.2 |
| 2: Scattered fibroglandular densities | 18,543 | 38.9 | 13,699 | 48.3 | 4,119 | 55.0 | 18,020 | 34.6 | 9,511 | 41.0 | 4,786 | 50.1 |
| 3: Heterogeneously dense | 24,833 | 52.2 | 11,865 | 41.8 | 2,643 | 35.3 | 24,547 | 47.2 | 9,566 | 41.2 | 3,383 | 35.4 |
| 4: Extremely dense | 2,817 | 5.9 | 1,510 | 5.3 | 445 | 5.9 | 6,258 | 12.0 | 2,607 | 11.2 | 882 | 9.2 |
| <i>Time since previous mammography**</i> | | | | | | | | | | | | |
| No previous mammography | 3,425 | 7.2 | 1,810 | 6.4 | 599 | 8.0 | 3,205 | 6.2 | 1,861 | 8.0 | 1,012 | 10.6 |
| < 12 months | 4,214 | 8.9 | 1,913 | 6.7 | 753 | 10.1 | 17,305 | 33.2 | 8,318 | 35.9 | 2,618 | 27.4 |
| 12 – 18 months | 22,726 | 47.7 | 16,574 | 58.4 | 3,725 | 49.8 | 17,258 | 33.2 | 6,692 | 28.9 | 2,845 | 29.8 |
| 19 – 30 months | 11,373 | 23.9 | 4,515 | 15.9 | 1,215 | 16.2 | 7,696 | 14.8 | 3,171 | 13.7 | 1,412 | 14.8 |
| 31 – 42 months | 2,653 | 5.6 | 1,662 | 5.9 | 498 | 6.7 | 2,691 | 5.2 | 1,309 | 5.6 | 616 | 6.5 |
| 43+ months | 3,218 | 6.8 | 1,898 | 6.7 | 693 | 9.3 | 3,895 | 7.5 | 1,842 | 7.9 | 1,041 | 10.9 |
| <i>BI-RADS assessment for all mammography</i> | | | | | | | | | | | | |
| 1: Negative | 13,978 | 29.4 | 9,509 | 33.5 | 2,805 | 37.5 | 23,201 | 44.6 | 10,551 | 45.5 | 4,932 | 51.7 |
| 2: Benign finding | 13,774 | 28.9 | 7,301 | 25.7 | 1,999 | 26.7 | 20,219 | 38.8 | 7,698 | 33.2 | 2,774 | 29.1 |

| | Additional evaluation of recent abnormal screening result | | | | | | Evaluation of breast problem | | | | | |
|---|---|------|-------|--|------------------------------------|------|------------------------------|-----|-------|--|------------------------------------|-----|
| | Non-vulnerable (score = 0) | | | Composite vulnerability score* | | | Non-vulnerable (score = 0) | | | Composite vulnerability score* | | |
| | N | % | | Moderately vulnerable (score = 1 or 2) | Highly vulnerable (score = 3 or 4) | | N | % | | Moderately vulnerable (score = 1 or 2) | Highly vulnerable (score = 3 or 4) | |
| 3.0: Probably benign finding/Need additional imaging [‡] | 13,614 | 28.6 | 7,396 | 26.1 | 1,564 | 20.9 | 3,948 | 7.6 | 2,278 | 9.8 | 748 | 7.8 |
| 3.0: Probably benign finding/Need additional imaging [‡] | 145 | 0.3 | 226 | 0.8 | 82 | 1.1 | 368 | 0.7 | 338 | 1.5 | 230 | 2.4 |
| 4: Suspiciously abnormal | 5,610 | 11.8 | 3,714 | 13.1 | 981 | 13.1 | 3,474 | 6.7 | 1,857 | 8.0 | 662 | 6.9 |
| 5: Highly suggestive of malignancy | 488 | 1.0 | 226 | 0.8 | 52 | 0.7 | 840 | 1.6 | 471 | 2.0 | 198 | 2.1 |

* Composite vulnerability scores were calculated by adding 1 for each of the following vulnerability indices met: *Education*: < 83% of the facility population completed high school; *Race/ethnicity*: > 30% of the facility population are minorities; *Rural/urban residence*: average percentage of rural residences among the facility population is > 52%; and *Income*: average median household income among the facility population is < \$45,000.

** For diagnostic mammography taken for 'Additional evaluation of a recent abnormal screening result' time since previous mammography is measured from the screen which led to the diagnostic mammography rather than the diagnostic mammography itself. For diagnostic mammography taken for 'Evaluation of a breast problem', time since previous mammography is measured from the diagnostic mammography.

[‡] Include a final recommendation for short interval follow-up, other work-up, additional imaging, clinical follow-up, or routine care

[‡] Included a final recommendation for biopsy, FNA, or surgical consult.

Table 2

Unadjusted sensitivity, false positive rate, and cancer detection rate for diagnostic mammography by vulnerability index of the facility at which the mammography was performed, stratified by radiologist indication

| | Additional evaluation of recent abnormal screening result | | | | | Evaluation of breast problem | | | | |
|--|---|------|-----------------|-------------------------|----------------------|------------------------------|------|-----------------|-------------------------|-----------------------|
| | N | % | Sensitivity (%) | False positive rate (%) | CDR (per 1000 exams) | N | % | Sensitivity (%) | False positive rate (%) | CDR* (per 1000 exams) |
| <i>Number of facilities</i> | 153 | | | | | 176 | | | | |
| <i>Education</i> | | | | | | | | | | |
| Non-vulnerable | 115 | 75.2 | 88.9 | 10.9 | 32.0 | 134 | 75.7 | 84.5 | 6.3 | 37.0 |
| Vulnerable | 38 | 24.8 | 87.9 | 12.0 | 32.8 | 42 | 23.7 | 82.5 | 8.0 | 36.9 |
| <i>Race/ethnicity</i> | | | | | | | | | | |
| Non-vulnerable | 127 | 83.0 | 88.9 | 10.8 | 32.3 | 146 | 82.5 | 84.4 | 6.4 | 36.9 |
| Vulnerable | 26 | 17.0 | 87.5 | 12.5 | 30.8 | 30 | 16.9 | 82.5 | 7.7 | 37.6 |
| <i>Rural/urban residence</i> | | | | | | | | | | |
| Non-vulnerable | 97 | 63.4 | 89.3 | 10.6 | 33.3 | 119 | 67.2 | 83.9 | 5.9 | 34.9 |
| Vulnerable | 56 | 36.6 | 87.3 | 12.0 | 29.4 | 57 | 32.2 | 84.9 | 8.5 | 43.5 |
| <i>Income</i> | | | | | | | | | | |
| Non-vulnerable | 98 | 64.1 | 89.8 | 10.5 | 32.4 | 115 | 65.0 | 84.7 | 6.4 | 36.2 |
| Vulnerable | 55 | 35.9 | 85.0 | 12.8 | 31.2 | 61 | 34.5 | 82.7 | 7.1 | 39.7 |
| <i>Composite vulnerability score**</i> | | | | | | | | | | |
| 0 | 62 | 40.5 | 90.2 | 10.1 | 34.4 | 77 | 43.5 | 84.7 | 5.7 | 35.2 |
| 1 | 42 | 27.5 | 87.2 | 12.0 | 26.6 | 44 | 24.9 | 83.7 | 8.0 | 39.6 |
| 2 | 21 | 13.7 | 85.0 | 13.1 | 32.0 | 26 | 14.7 | 83.9 | 7.8 | 39.8 |
| 3 | 21 | 13.7 | 87.0 | 12.1 | 31.5 | 22 | 12.4 | 81.8 | 7.7 | 38.9 |
| 4 | 7 | 4.6 | 92.9 | 11.2 | 52.7 | 7 | 4.0 | 88.5 | 8.6 | 48.4 |

* CDR = Cancer detection rate

Table 3a

Adjusted sensitivity, false positive rates, and cancer detection rate for diagnostic mammography by vulnerability index of the facility at which the mammography was performed*

| Vulnerability Index | Additional evaluation of recent abnormal screening result | | | | | | | | | | | |
|---|---|----------------|-------------------|---------------------------|----------------|---------------------|---------------------------|----------------|-------------------|-----------------|-------------------------|--|
| | Sensitivity | | | | | False positive rate | | | | | CDR** (per 1,000 exams) | |
| | Non-vulnerable (%) | Vulnerable (%) | OR (95% CI)*** | Non-vulnerable (%) | Vulnerable (%) | OR (95% CI)*** | Vulnerable (%) | Non-vulnerable | Vulnerable | OR (95% CI)**** | OR (95% CI)**** | |
| <i>Education</i> | 87.8 | 84.7 | 0.77 (0.44, 1.34) | 10.3 | 10.1 | 0.97 (0.76, 1.24) | 30.3 | 27.1 | 0.89 (0.69, 1.15) | | | |
| <i>Race/ethnicity</i> | 87.8 | 83.5 | 0.70 (0.37, 1.31) | 10.3 | 10.4 | 1.01 (0.76, 1.35) | 30.4 | 26.0 | 0.85 (0.63, 1.15) | | | |
| <i>Rural/urban residence</i> | 87.7 | 87.2 | 0.96 (0.61, 1.50) | 9.8 | 11.0 | 1.14 (0.94, 1.38) | 28.4 | 32.5 | 1.15 (0.94, 1.41) | | | |
| <i>Income</i> | 88.6 | 83.4 | 0.64 (0.40, 1.02) | 9.6 | 12.8 | 1.39 (1.13, 1.70) | 30.5 | 27.6 | 0.90 (0.72, 1.12) | | | |
| <i>Composite Vulnerability Score</i> [†] | | | | | | | | | | | | |
| <i>0</i> | Reference | | 89.4 | Reference | | 9.4 | Reference | | 31.6 | | | |
| <i>1</i> | 0.72 (0.45, 1.15) | | 86.0 | 1.11 (0.91, 1.37) | | 10.4 | 0.87 (0.70, 1.08) | | 27.7 | | | |
| <i>2</i> | 0.56 (0.32, 0.98) | | 82.7 | 1.32 (1.02, 1.70) | | 12.0 | 0.88 (0.68, 1.16) | | 28.1 | | | |
| <i>3</i> | 0.62 (0.33, 1.19) | | 84.2 | 1.30 (0.96, 1.76) | | 11.9 | 0.87 (0.64, 1.19) | | 27.7 | | | |
| <i>4</i> | 1.64 (0.34, 7.97) | | 93.2 | 1.02 (0.63, 1.66) | | 9.6 | 1.42 (0.84, 2.40) | | 44.1 | | | |
| | <i>p for trend = 0.61</i> | | | <i>p for trend = 0.72</i> | | | <i>p for trend = 0.22</i> | | | | | |

* Adjusted for mammography-level registry, age, time since previous mammography, and BI-RADS breast density.

** CDR = Cancer detection rate

*** Non-vulnerable is taken to be the reference group.

† Composite vulnerability scores were calculated by adding 1 for each of the vulnerability indices met.

Table 3b

Adjusted sensitivity, false positive rate, and cancer detection rate for diagnostic mammography by vulnerability index of the facility at which the mammography was performed*

| | Evaluation of breast problem | | | | | | | | | |
|--------------------------------------|------------------------------|----------------|-------------------|---------------------------|----------------|-------------------|-------------------------|------------|-------------------|---------------------------|
| | Sensitivity | | | False positive rate | | | CDR** (per 1,000 exams) | | | |
| | Non-vulnerable (%) | Vulnerable (%) | OR (95% CI)*** | Non-vulnerable (%) | Vulnerable (%) | OR (95% CI)*** | Non-vulnerable | Vulnerable | OR (95% CI)*** | OR (95% CI)*** |
| Vulnerability Index | | | | | | | | | | |
| <i>Education</i> | 81.9 | 79.8 | 0.87 (0.63, 1.20) | 6.1 | 8.2 | 1.39 (1.08, 1.79) | 36.5 | 39.6 | 1.09 (0.88, 1.35) | |
| <i>Race/ethnicity</i> | 81.5 | 81.1 | 0.97 (0.67, 1.42) | 6.2 | 8.0 | 1.32 (0.98, 1.76) | 35.9 | 44.3 | 1.25 (0.98, 1.60) | |
| <i>Rural/urban residence</i> | 80.4 | 83.5 | 1.24 (0.93, 1.66) | 5.7 | 8.5 | 1.55 (1.27, 1.88) | 34.6 | 42.9 | 1.26 (1.05, 1.50) | |
| <i>Income</i> | 81.6 | 80.9 | 0.95 (0.70, 1.29) | 6.0 | 7.8 | 1.34 (1.08, 1.66) | 35.6 | 41.0 | 1.16 (0.97, 1.40) | |
| Composite Vulnerability Score | | | | | | | | | | CDR |
| <i>0</i> | Reference | Reference | 81.2 | Reference | Reference | 5.4 | Reference | Reference | Reference | 33.5 |
| <i>1</i> | 1.10 (0.78, 1.55) | | 82.5 | 1.31 (1.06, 1.63) | | 6.9 | 1.15 (0.94, 1.41) | | | 38.3 |
| <i>2</i> | 1.01 (0.69, 1.48) | | 81.3 | 1.39 (1.08, 1.78) | | 7.3 | 1.25 (0.99, 1.57) | | | 41.2 |
| <i>3</i> | 0.89 (0.60, 1.32) | | 79.4 | 1.71 (1.27, 2.32) | | 8.8 | 1.28 (0.98, 1.67) | | | 42.2 |
| <i>4</i> | 1.72 (0.70, 4.20) | | 87.9 | 2.12 (1.36, 3.32) | | 10.6 | 1.40 (0.91, 2.14) | | | 45.9 |
| | <i>p for trend</i> = 0.36 | | | <i>p for trend</i> < 0.01 | | | | | | <i>p for trend</i> = 0.10 |

* Adjusted for mammography-level registry, age (5-year bands), time since previous mammography, and BI-RADS breast density.

** CDR = Cancer detection rate

*** Non-vulnerable is taken to be the reference group.

† Composite vulnerability scores were calculated by adding 1 for each of the vulnerability indices met.