Published in final edited form as:

J Clin Epidemiol. 2010 April ; 63(4): 441–451. doi:10.1016/j.jclinepi.2009.06.008.

Decreased Accuracy In Interpretation Of Community-Based Screening Mammography For Women With Multiple Clinical Risk Factors

Andrea J. Cook, PhD^{1,2}, Joann G. Elmore, MD, MPH³, Diana L. Miglioretti, PhD^{1,2}, Edward A. Sickles, MD⁴, Erin J. Aiello Bowles, MPH¹, Gary R. Cutter, PhD⁵, and Patricia A. Carney, PhD⁶

¹Group Health Center for Health Studies, Group Health Cooperative, Seattle, WA

²Department of Biostatistics, University of Washington School of Public Health and Community Medicine, Seattle, WA

³Department of Medicine, University of Washington School of Medicine, Harborview Medical Center, Seattle, WA

⁴Department of Radiology, University of California School of Medicine, UCSF Medical Center, San Francisco, CA

⁵Department of Biostatistics, University of Alabama School of Public Health, Birmingham, AB

⁶Department of Family Medicine and of Public Health and Preventive Medicine, Oregon Health & Science University, Portland, OR

Abstract

Objective—To assess the impact of women's breast cancer risk factors (use of hormone therapy, family history of breast cancer, previous breast biopsy) on radiologists' mammographic interpretive performance and whether the influence of risk factors varies according to radiologist characteristics.

Study Design and Setting—Screening mammograms (n=638,947) performed from 1996 to 2005 by 134 radiologists from three Breast Cancer Surveillance Consortium registries were linked to cancer outcomes, radiologist surveys, and patient questionnaires. Interpretive performance measures were modeled using marginal and conditional logistic regression.

Results—Having one or more clinical risk factors was associated with higher recall rates [1 vs. 0 risk factors: odds ratio=1.17, 95% confidence interval=(1.15-1.19); ≥2 vs. 0: 1.43(1.40-1.47)] and lower specificity [1 vs. 0: 0.86(0.84-0.88); ≥2 vs. 0: 0.70(0.68-0.72)] without a corresponding improvement in sensitivity and only a small increase in positive predictive value [1 vs. 0: 1.08 (0.99-1.19); ≥2 vs. 0: 1.12(0.99-1.26)]. There was no indication that influence of risk factors varied by radiologist characteristics.

Conclusion—Women with clinical risk factors who undergo screening mammography are more likely recalled for false positive evaluation without an associated increase in cancer detection.

© 2009 Elsevier Inc. All rights reserved

Address for correspondence: Andrea J. Cook, PhD Group Health Center for Health Studies 1730 Minor Avenue, Suite 1600, Seattle, WA 98101 Phone (206) 287-4257 Fax (206) 287-4710 cook.aj@ghc.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Radiologists and patients with risk factors should be aware of this increased risk of adverse screening events.

Keywords

Breast cancer screening; mammography; radiologist performance; hormone replacement therapy; family history; breast biopsy

What is New?

Women with clinical risk factors for breast cancer who obtain a screening mammogram are more likely to be recalled for additional evaluation and biopsied for non-cancerous findings with no associated increase in the probability of detecting cancer. The independent effect of individual breast cancer risk factors have previously been reported, but the effect of multiple risk factors on the interpretation of mammography has not been shown. This information should be conveyed to radiologists and high-risk patients so they are aware of the increased likelihood of being recalled for benign findings during a screening mammogram.

Introduction

In mammography practice, a patient's clinical history provides important information about breast cancer risk. Interest in the role of risk factors on cancer screening guidelines is growing (1), but whether interpretive performance could be improved by using patient risk factor information during mammogram interpretation has not been adequately studied in community-based mammography practice or has produced conflicting results. A study using mammographic test sets suggested that knowledge of patient clinical risk factors alter a radiologist's level of diagnostic suspicion without improving performance (2).

In contrast, two cross-sectional observational studies found that while sensitivity was not different for women with a first-degree relative with breast cancer, positive predictive value was higher among women with a family history compared to those without (3,4).

The goal of this study was to assess how the existence of breast cancer risk factors may be associated with interpretive performance and whether any relationships between risk and performance depended on characteristics of the interpreting radiologist. We might expect radiologists to have a lower threshold for recalling women with multiple risk factors for further work-up, which would theoretically result in finding more cancers, but might also increase false-positive rates for these high-risk women. If patient risk factor information improved interpretation performance, we would expect an equal or higher sensitivity for detecting cancer among high-risk women without an associated increase in the false-positive rate when these measures are compared to women without risk factors. Alternatively, we might expect women without risk factors to be less likely to have a false-positive examination without a corresponding reduction in the sensitivity for detecting cancer. Using data collected by three mammography registries that participate in the Breast Cancer Surveillance Consortium, we had the unique opportunity to study how radiologists' performance may be associated with women's risk.

Methods

Study population

Three mammography registries that are part of the National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC; <http://breastscreening.cancer.gov>) contributed data for this study: Group Health, a non-profit integrated healthcare organization in the Puget Sound region of Washington state; the New Hampshire Mammography Network, which captured approximately 90% of mammograms performed in New Hampshire; and the Colorado Mammography Program, which captured approximately 50% of mammograms performed in the Denver metropolitan area during the study time period (5). These registries collect patient demographic and clinical information each time a woman receives a mammography examination at a participating facility. This information is linked to regional cancer registries and pathology databases to determine cancer outcomes. Data from the registries were pooled at the BCSC Statistical Coordinating Center for analysis.

Each registry and the SCC 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 SCC have received a Federal Certificate of Confidentiality and other protection for the identities of women, physicians, and facilities that are subjects of this research.(6)

Radiologists who interpreted mammograms at a facility contributing to any of the three registries between January 1996 and December 2001 were invited to participate in a mailed survey in 2002, using survey methods previously described (7). Of the 139 radiologists who responded to the survey (77% response rate), we excluded those who had no screening mammograms in the database during the study years ($N=4$) or were missing information on radiologist variables previously found to be associated with interpretive performance (years of experience and self-reported number of mammograms interpreted; $N=1$). Twenty facilities, from the initial 121, were excluded from these analyses due to under-reporting of patient risk information [current use of hormone therapy (HT), any family history of breast cancer, and previous biopsy] on $>75\%$ of a single clinical risk factor or not reporting any of the clinical risk factor information $\geq 90\%$ of the time. There were 356 mammograms excluded from the analysis due to being conducted at one of the twenty facilities. The final study sample included 134 radiologists from 101 facilities.

We limited this study to bilateral mammograms indicated as screening by the interpreting radiologist and performed between January 1, 1996 and October 31, 2005. Mammograms interpreted through October 31, 2005 were included to allow adequate time for ascertainment of cancers diagnosed within 365 days of the mammogram. We excluded mammograms performed on women who had a breast imaging examination within the prior 9 months and mammograms performed on women who indicated nipple discharge or breast lump at the time of the mammogram ($N=12,153$) to avoid misclassifying diagnostic mammograms as screening. We excluded mammograms performed on women with breast augmentation, reconstruction, reduction, or mastectomy, women <30 or >90 years old, and women with a prior history of breast cancer ($N=2,460$). We removed mammograms without information on the woman's breast density, age, or time since last mammogram ($N=129,045$), because these were deemed important potential patient-level confounders associated with interpretive performance (8,9). We then excluded mammograms missing all three risk factors of interest: current HT use, any family history of breast cancer, and previous breast biopsy ($N=20,853$). We removed these mammograms because we could not presume that if all of these risk factors were missing that the data was not available to the radiologist at the time of screening, but it could equally likely be just not entered into the data system. The cancer rate of those excluded mammograms were

elevated in this subpopulation compared to the final analysis dataset (8.4 per 1000 compared 4.7 per 1000), but the recommendation for biopsy rate did not differ (1.7 per 100 compared to 1.4 per 100) between these excluded mammograms and mammograms included in our final analysis dataset. Our analysis included the remaining 638,947 screening mammograms, which were interpreted by the 134 radiologists described above.

Measurements

Data on radiologist characteristics were obtained from the previously described self-administered mailed survey, which included questions about demographic characteristics, experience, and clinical practice characteristics in the prior year. We examined several radiologist characteristics previously found to be associated with the accuracy of screening mammography (7): radiologist age, years of experience, self-reported number of mammograms interpreted in the prior year, and affiliation with an academic institution.

Patient risk factor information collected at the time of the mammogram included age, time since last mammogram, current HT use, ever having a breast biopsy and family history of breast cancer. We derived a clinical history risk factor score using three risk factors available to the radiologist at the time of the mammogram: current HT use (Yes, No), family history of breast cancer (Yes, No), and ever having a breast biopsy (Yes, No). If at least one of the three clinical risk factors were available then we treated any missing clinical risk factors as *not* being available to the radiologist at the time of screening and therefore contributing 0 to the overall clinical risk factor score which ranged from 0 to 3.

We also obtained the radiologists' Breast Imaging Reporting and Data System (BI-RADS®) assessment(10) and recommendation, and BI-RADS® mammographic breast density category (1: entirely fatty, 2: scattered fibroglandular tissue, 3: heterogeneously dense, and 4: extremely dense). Because breast density is a radiologist-defined variable that represents how the radiologist estimates the percentage of mammographically dense tissue and may vary from radiologist to radiologist, we did not include it in the patient risk factors studied. Further detail about the BCSC and the data they collect can be obtained from <http://breastscreening.cancer.gov/>.

Mammography examinations given an initial BI-RADS® assessment of 0 (need additional imaging), 4 (suspicious abnormality), or 5 (highly suggestive of cancer) were considered to be *positive* (10). An initial BI-RADS® assessment of 3 (probably benign) with a recommendation for immediate follow-up was also considered *positive*. We classified as *negative* those mammograms given an initial BI-RADS® assessment of 1 (negative), 2 (benign), or 3 (probably benign) without a recommendation for immediate follow-up. *Recall rate* was defined as the percentage of positive examinations among all screening mammograms. *Biopsy recommendation rate* was defined as the percentage of women who were given a final BI-RADS assessment of 4 or 5 or an assessment of 0 or 3 with a recommendation for a biopsy at the end of imaging work-up and within 90 days of the initial screening examination among all screening mammograms.

We linked women with tumor registry data to determine whether they were diagnosed with invasive breast cancer or ductal carcinoma *in situ* (DCIS) within 1 year of the mammography examination and prior to the next screening mammogram. We considered a mammogram given a positive assessment to be a *true positive* if breast cancer was diagnosed within the follow-up period. We considered a mammogram given a negative assessment to be a *true negative* if breast cancer was not diagnosed within the follow-up period. We define the *cancer rate* as the percentage breast cancers among all screening mammograms. Note that this rate differs from the cancer rate that would occur in a population-based sample as our sample only included women undergoing screening. *Sensitivity* was defined as the percentage of true-positive

examinations among women diagnosed with breast cancer. *Specificity* was defined as the percentage of true-negative examinations among women without a breast cancer diagnosis. *Positive predictive value* (PPV) was defined as the percentage of true positive examinations among women with positive examinations.

Statistical Analysis

We calculated unadjusted rates of cancer diagnoses per 1,000 screening mammograms, recall rates and biopsy recommendations per 100 screening mammograms, sensitivity per 100 mammograms on women *with* cancer, specificity per 100 mammograms on women *without* cancer, and PPV per 100 *positive* screening mammograms with 95% confidence intervals (CIs) separately for each radiologist and by the risk factors of the women and characteristics of the radiologists. To assess the relationship between a radiologist characteristics and interpretive performance, we fit marginal logistic regression models using generalized estimating equations (GEE) with a sandwich variance estimate after assuming independent correlation structure to account for potential correlation among multiple mammograms interpreted by the same radiologist and performed on the same woman (11-13).

For the analysis of recall, we included all women and modeled the probability of a positive mammogram (versus negative mammogram) as the outcome in the logistic regression models. For the analysis of sensitivity, we included women with a diagnosis of cancer in the follow-up period and modeled the probability of a true-positive mammogram (versus a false-negative mammogram) as the outcome. For specificity, we included women without a diagnosis of cancer and modeled the probability of a true-negative mammogram (versus a false-positive mammogram) as the outcome. For PPV, we included women with a positive mammogram and modeled the probability of a true-positive mammogram (versus false-positive mammogram) as the outcome.

We also fit marginal logistic regression models using GEE for recall rate, sensitivity, specificity, and PPV to assess interactions between the effect of radiologist characteristics and patient clinical risk factor score (0, 1, or ≥ 2) available to the radiologist at the time of the mammography examination adjusting for patient age (linear effect), mammographic breast density, and BCSC site. Patient risk factor score values 2 and 3 were collapsed into one category because very few women had all 3 risk factors under study.

For the multivariable analyses we assessed the relationship between performance measures and the effect of varying woman-level characteristics within radiologists using conditional logistic regression stratified on radiologist. Conditional logistic regression uses the statistical technique of conditioning to remove the effects of any heterogeneity among radiologists(14). By using this technique we assessed if there was a difference in an individual radiologist's mammography performance related to the number of patient's clinical risk factors, controlling for the effects of any radiologist-level characteristics such as radiologist experience. Importantly, this approach conditions out any effects due to differences in case-mix across radiologists (e.g., radiologists who primarily interpret mammograms for women at high risk of breast cancer may have a different mammography performance compared to radiologists who see women with fewer risk factors). In this way, conditional logistic regression estimates associations solely due to varying covariates such as a patient's number of clinical risk factors within radiologists. Patient factor score, linear effect of patient age, time since last mammogram, and mammographic breast density were all included in the model.

To assess if each clinical risk factor contributed similarly to the risk factor score, we ran multivariate analyses and included an indicator for each risk factor combination (i.e. no risk factors, current HT use only, family history only, benign biopsy only, current HT use and family history only, current HT use and benign biopsy, family history and benign biopsy, and all three

risk factors) into a single model adjusted for linear effect of patient age, time since last mammogram, and mammographic breast density. This allowed for each individual risk factor combination to be estimated instead of the risk factor score.

We calculated the expected PPV for the clinical risk factor score groups given the observed cancer rates and a constant sensitivity and specificity across groups using the following formula:

$$\frac{\text{Sensitivity} * \text{CancerRate}}{\text{Sensitivity} * \text{CancerRate} + (1 - \text{Specificity}) * (1 - \text{CancerRate})}$$

The corresponding expected PPV Odds Ratio comparing Group B (e.g., high risk women) to Group A (e.g., low risk women) then becomes:

$$\frac{\text{CancerRate}_B / (1 - \text{CancerRate}_B)}{\text{CancerRate}_A / (1 - \text{CancerRate}_A)}$$

which does not depend on the value of the sensitivity or specificity when these measures are equal for the two groups. Therefore, if high risk women have a higher cancer rate, they would have a higher expected PPV even if they do not have improved sensitivity or specificity. We compared the expected and observed PPV odds ratios to evaluate whether any differences in PPV could be explained simply by differences in cancer rates versus differences in interpretive performance. See the appendix for derivation of the expected PPV and the PPV Odds Ratio.

Data analyses were conducted using SAS® software, Version 9.1 (SAS institute, Cary, NC). *P*-values are two-sided, and we only included covariates in the model that were significant at the 0.10 level in the univariable models.

Results

As expected, the cancer rate increased with increasing patient age and breast density and was higher among women with each of the three clinical risk factors of interest (Table 1). There were 314,497 (49.2%) mammograms performed on women with no clinical risk factors, 247,759 (38.8%) on women with one, 69,426 (10.9%) on women with two, and 7,265 (1.1%) on women with all three risk factors. Among mammography exams on women with only one clinical risk factor 127,026 (51.3%) were current HT users, 60,362 (24.4%) had a family history of breast cancer, and 60,371 (24.4%) had a previous biopsy. Among mammography exams on women with two clinical risk factors 19,493 (28.1%) were both current HT users and had a family history, 33,937 (48.9%) were both current HT users and had a previous biopsy, and 15,996 (23.0%) had both a family history and a previous biopsy.

The cancer rate was 3.8 per 1,000 screens for women without any known clinical risk factors at the time of screening and 10.2 per 1,000 screens for women with current HT use, family history, and previous biopsy. Recall rate increased from 9.8 to 13.2 and the percentage of exams with a radiologists' recommendation to biopsy doubled from 1.2 to 2.2 for women with all three risk factors compared to women without any known risk factors at the time of screening.

Interpretive performance by radiologist characteristics is shown in Table 2 stratified by the risk factor scores of the women whose mammograms they interpreted. Across all patient risk scores, recall rate was significantly higher among younger radiologists ($p=0.05$), those who interpreted more mammograms per year ($p<0.01$), and those that had their primary affiliation with an academic institution ($p<0.01$), but did not vary significantly by those who had different years

of experience interpreting mammography ($p=0.21$). The recall rate was more likely to be higher for radiologists who interpreted mammograms on higher-risk women compared to radiologists who interpreted mammograms on low-risk women ($p<0.01$). The magnitude of the risk factor effect did not differ significantly within any radiologist characteristic subgroups ($p>0.10$ in all cases). PPV was not significantly higher for any radiologist characteristics, but PPV increased with increased patient risk factor score.

Multivariable results for the recall rate, sensitivity, specificity, and PPV, adjusting for radiologist and patient characteristics are presented in Table 3. Recall rate significantly increased with higher patient risk factor score [ORs 1 vs. 0 risk factors: 1.17 (95% CI 1.15-1.19), ≥ 2 vs. 0: 1.43 (95% CI 1.40-1.47)]. The risk factor score had no effect on sensitivity. Specificity decreased with increasing risk factor score [ORs 1 vs. 0 risk factors: 0.86 (95% CI 0.84-0.88), ≥ 2 vs. 0: 0.70 (95% CI 0.68-0.72)]. PPV increased slightly with increasing risk factor score, but the statistical significance of this association was borderline [ORs 1 vs. 0: 1.08 (95% CI 0.99-1.19), ≥ 2 vs. 0: 1.12 (95% CI 0.99-1.26)].

The extent of a patient's breast density also had a significant association with performance within an individual radiologist after adjusting for all other patient risk factors (Table 3). Women with less dense breasts were less likely to be recalled and had higher sensitivity and specificity compared to women with denser breast tissue. PPV did not change with breast density. With increasing time since last mammogram, recall rate and sensitivity increased while specificity decreased. PPV was higher for women who had their last mammogram 3-4 years previously compared to ≤ 2 years.

Further multivariable results assessing each potential combination of risk factors, adjusting for radiologist and patient characteristics are presented in Table 4. Overall, either being a current HT user or having a previous biopsy significantly increased recall rate compared to having no risk factors, but having a family history alone did not influence recall rate after adjusting for HT use and previous biopsy. No obvious differences in sensitivity for the three risk factors were observed. Current HT use and previous biopsy accounted for most of the observed decrease in specificity. Having a family history improved PPV, while current HT use or a previous biopsy had no influence.

PPV did not increase with increasing clinical risk factor score to the magnitude that is expected based on the observed cancer rates (Table 5). If sensitivity and specificity were not influenced by the presence of clinical risk factors (e.g., were fixed across all risk factor scores), the odds of cancer given a positive mammogram (PPV) would be expected to increase by 1.39 times for women with one clinical risk factor compared to no risk factors and 1.92 times comparing women with two or more risk factors to no clinical risk factors. We observed only an estimated OR of 1.08 (95% CI 0.99-1.19) and 1.12 (95% CI 0.99-1.26), respectively.

Discussion

Women with risk factors for breast cancer were more likely to be recalled for additional imaging and biopsy after screening mammography when they did not have breast cancer. Unfortunately, this increase in recall did not result in an increased probability of detecting cancer when it was present. Moreover, although the PPV trended towards being higher among women with risk factors, the magnitude of this increase was not as large as we would expect due to the actual increased cancer rate noted among these women.

Our findings suggest that radiologists' interpretations may be influenced by information about patient risk available to them at the time of the mammogram. Availability of risk factor information may improve performance among women with no risk factors, as these women were less likely to have a false-positive examination with no change in the probability of

detecting cancer. False-positive rates for low risk women are within the recommended U.S. guidelines of <10% (10). In contrast, for every 1,000 women with three risk factors screened, 34 more women were recalled for additional imaging compared to those with no risk factors, and 10 more women were recommended for biopsy. Unfortunately, this increase in work-up did not correspond to an increase in detecting cancer when it was present.

It is possible that radiologists are overly cautious about interpreting mammographic findings among women with risk factors, because they overestimate these women's pre-test probability of cancer. Indeed, in one study of these same radiologists, 96% overestimated an individual woman's 5- year risk of a diagnosis of breast cancer, particularly those with risk factors such as family history of breast cancer and prior biopsy (15). This suggests that radiologists may be over influenced by some risk factors when they are deciding whether or not to recall women for additional work up or are considering whether or not to biopsy a possible abnormality.

While observational studies have shown that mammography performed on women currently using postmenopausal hormone therapy (HT) have higher false positive rates (8,15-18) and lower sensitivity (8,18) it is not clear if this is completely caused by HT's effect on breast density or if radiologists are being influenced by the knowledge of HT use, or both. In our study, we adjusted our analyses for radiologists' interpretation of breast density. Thus, our findings should not be affected by breast density patterns. The increased recall and biopsy recommendation rates may be appropriate if true abnormalities are visually apparent on the exam.

It is possible that women with risk factors have different breast architecture and lesions visible on their exams compared to women with no risk factors. For example, women taking conjugated equine estrogen have been shown to be at increased risk of benign proliferative breast disease (16) and women with a family history of breast cancer were shown to have an increased risk of biopsy confirmed benign breast disease (17,18). In another study, Berg and colleagues found no increased risk for breast cancer associated with radial scar beyond that of proliferative disease without atypia that occurs in the general population (23). Another study (19) examined multiple benign breast pathology (non-proliferative, proliferative, or proliferative with atypia) to address the issue of the contribution of concurrent multiple lesions on breast cancer risk. They found that >70% had more than one type of benign lesion and that multiple benign lesions and patient age were associated with increased risk of subsequent breast cancer. In yet another study(25), investigators found that among women with atypical hyperplasia, multiple foci of atypia and presence of histologic calcifications appears to indicate very high risk (> 50% risk at 20 years), but that a positive family history does not further increase risk in women with atypia. These studies suggest that visual distortions possibly associated with risk are likely present and may need work-up. However, more research is needed to determine how best to provide surveillance to women with multiple risk factors for breast cancer.

We know of only a few studies that have examined the influence of clinical information on the accuracy of mammography, and most of these focused on diagnostic rather than screening mammography (2,20,21) and have produced conflicting results (20,21). Prior studies that evaluated the influence of risk factors on interpretive performance had several weaknesses, including the small number of radiologists studied (≤ 10). Prior studies also measured performance using test sets, which may not replicate actual practice(2,22,23)

Our study had several strengths in contrast to the published literature. We examined mammography performance data from actual clinical practice in three geographically distinct regions in the US as opposed to evaluating performance using test sets of mammograms. In addition, because of the large number of radiologists and mammograms included, our findings

are likely more stable than results from test sets. Another important strength of our study is that we had the ability to perform analyses that accounted for radiologists' characteristics previously shown to influence performance in other studies, such as radiologist age, number of years interpreting mammograms, and annual interpretive volume (7, 28). We, therefore, had the unique opportunity to conduct within-radiologist comparisons by applying conditional logistic regression models. This statistical technique can be used to account for all between radiologist differences including differences in case-mix (14). Results are therefore interpreted on the mammogram level and measure the effect of changing the number of clinical risk factors for an individual radiologist's performance, removing the potential effect of between radiologist differences. This is a key improvement from most analyses conducted in this area, since previous applied methods did not separate the between and within-in radiologist effects. Between radiologist differences may overshadow any within-in radiologist relationships that are of actual interest. For example, the between effect for changing number of clinical risk factors would be estimating whether radiologists who interpret mammograms more often on higher risk women have a different interpretative performance than those who typically interpret mammograms on lower risk women. This could occur if women with more risk factors were more likely to obtain mammograms at particular facilities, such as academic facilities, which have different baseline interpretive performance rates (24). One could attempt to control for confounders through adjustment, but it is difficult to identify and accurately measure all possible confounders. Using conditional logistic regression assures that case-mix is being fully adjusted for without making such strong assumptions and therefore you are able to easily and robustly estimate pure within radiologist effects.

Our study also had some limitations. We did not examine whether the radiologists actually used the risk factor information at the time of the mammogram interpretation; rather we only looked at whether the risk factors were reported at the screening examination. A follow-up study should identify whether and how radiologists use clinical history in screening mammography interpretation, as this may be an area in which the performance can be improved. In addition, we cannot determine whether the increased recall and biopsy rates observed for high risk women were appropriate, given that these women may be at increased risk of benign breast disease that requires additional work-up to rule-out cancer.

In conclusion, our large multi-center study found that radiologists are less likely to recall women who have no risk factors for breast cancer, reducing the number of false-positive examinations in these women without increasing the probability of missing cancers. However, our findings also suggest that radiologists may be over using clinical history about women's risk factors in their recall of more women with risk factors without increasing the probability of detecting cancer, particularly for women who are currently using HT and/or have had a previous biopsy. Alternatively, this increased recall may be appropriate given that the clinical risk factors we studied are associated with an increased risk of benign breast disease. Women with multiple risk factors for breast cancer should be informed that they are at elevated risk of being recalled for additional follow-up and biopsy following a screening mammography examination even when breast cancer is not present.

Acknowledgments

This work was supported by the Agency for Healthcare Research and Quality (HS-10591) and the National Cancer Institute (1R01 CA107623; 1K05 CA104699; Breast Cancer Surveillance Consortium: U01 CA63731, U01 CA86082, U01 CA63736, and U01CA86076). The collection of cancer incidence data used in this study was supported, in part, by the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #U55/CCR921930-02 awarded to the Public Health Institute; Cancer Surveillance System of the Fred Hutchinson Cancer Research Center, which is funded by Contract No. N01-CN-67009 and N01-PC-35142 from the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute with additional support from the Fred Hutchinson Cancer Research Center and the State of Washington; New Hampshire State Cancer Registry supported in part by cooperative agreement U55/CCU-121912 awarded to the New Hampshire Department of Health

and Human Services, Division of Public Health Services, Bureau of Disease Control and Health Statistics, Health Statistics and Data Management Section; Colorado Central Cancer Registry, which is partially supported by the Colorado State General Fund and the federal Centers for Disease Control and Prevention (National Program of Cancer Registries) under Cooperative Agreement U58000848. The ideas and opinions expressed herein are those of the authors and endorsement by the New Hampshire Department of Health and Human Services; the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors is not intended nor should be inferred.

Appendix: Derivation of Expected PPV and PPV odds ratios

The positive predictive value (PPV) is measuring the probability of having cancer among women with a positive mammogram assessment. Using Bayes formula several times as shown below the components of PPV can be transformed to be dependent on the probability of cancer (cancer rate), the probability of positive mammogram assessment among women with cancer (sensitivity), and the probability of positive assessment among women without cancer (1-specificity).

$$\begin{aligned} PPV &= P(\text{Cancer}|\text{Positive}) = \frac{P(\text{Cancer and Positive})}{P(\text{Positive})} \\ &= \frac{P(\text{Positive}|\text{Cancer}) * P(\text{Cancer})}{P(\text{Positive}|\text{Cancer}) * P(\text{Cancer}) + P(\text{Positive}|\text{NoCancer}) * P(\text{NoCancer})} \end{aligned}$$

Therefore, the expected PPV based on cancer rate, sensitivity and specificity is

$$\text{Expected PPV} = \frac{\text{Sensitivity} * \text{CancerRate}}{\text{Sensitivity} * \text{CancerRate} + (1 - \text{Specificity}) * (1 - \text{CancerRate})}$$

The PPV odds ratio comparing Group B to Group A by definition maybe calculated as

$$\begin{aligned} OR &= \frac{PPV_B / (1 - PPV_B)}{PPV_A / (1 - PPV_A)} \\ &= \frac{P(\text{Cancer}|\text{Positive}, \text{Group B}) / P(\text{No Cancer}|\text{Positive}, \text{Group B})}{P(\text{Cancer}|\text{Positive}, \text{Group A}) / P(\text{No Cancer}|\text{Positive}, \text{Group A})} \\ &= \frac{P(\text{Positive}|\text{Cancer}, \text{Group B}) * P(\text{Cancer}|\text{Group B}) / (P(\text{Positive}|\text{NoCancer}, \text{Group B}) * P(\text{NoCancer}|\text{Group B}))}{P(\text{Positive}|\text{Cancer}, \text{Group A}) * P(\text{Cancer}|\text{Group A}) / (P(\text{Positive}|\text{NoCancer}, \text{Group A}) * P(\text{NoCancer}|\text{Group A}))} \cdot 1 \end{aligned}$$

If we assume sensitivity and specificity are the same in Groups A and B (i.e., $P(\text{Positive} | \text{Cancer}, \text{Group B}) = P(\text{Positive} | \text{Cancer}, \text{Group A})$ and $P(\text{Negative} | \text{No Cancer}, \text{Group B}) = P(\text{Negative} | \text{No Cancer}, \text{Group A})$) then the expected PPV OR is

$$\begin{aligned} \text{Expected PPV OR} &= \frac{P(\text{Cancer}|\text{Group B}) / (1 - P(\text{Cancer}|\text{Group B}))}{P(\text{Cancer}|\text{Group A}) / (1 - P(\text{Cancer}|\text{Group A}))} \\ &= \frac{\text{CancerRate}_B / (1 - \text{CancerRate}_B)}{\text{CancerRate}_A / (1 - \text{CancerRate}_A)} \end{aligned}$$

References

1. Elmore J, Choe J. Breast cancer screening for women in their 40s: moving from controversy about data to helping individual women. *Ann Intern Med* 2007;146(7):529–31. [PubMed: 17404356]
2. Elmore JG, Wells CK, Howard DH, Feinstein AR. The impact of clinical history on mammographic interpretations. *JAMA* 1997;277(1):49–52. [PubMed: 8980210]
3. Kerlikowske K, Carney PA, Geller B, Mandelson MT, Taplin SH, Malvin K, Ernster V, Urban N, Cutter G, Rosenberg R, Ballard-Barbash R. Performance of screening mammography among women

- with and without a first-degree relative with breast cancer. *Ann Intern Med* 2000;133(11):855–63. [PubMed: 11103055]
4. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA* 1996;276(1):33–8. [PubMed: 8667536]
 5. Ballard-Barbash R, Taplin SH, Yankaskas BC, Ernster VL, Rosenberg RD, Carney PA, Barlow WE, Geller BM, Kerlikowske K, Edwards BK, Lynch CF, Urban N, Chrvala CA, Key CR, Poplack SP, Worden JK, Kessler LG. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *AJR Am J Roentgenol* 1997;169(4):1001–8. [PubMed: 9308451]
 6. Carney PA, Geller BM, Moffett H, Ganger M, Sewell M, Barlow WE, Stalnaker N, Taplin SH, Sisk C, Ernster VL, Wilkie HA, Yankaskas B, Poplack SP, Urban N, West MM, Rosenberg RD, Michael S, Mercurio TD, Ballard-Barbash R. Current medicolegal and confidentiality issues in large, multicenter research programs. *Am J Epidemiol* 2000;152(4):371–8. [PubMed: 10968382]
 7. Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, Hendrick RE, Elmore JG. Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst* 2004;96(24):1840–50. [PubMed: 15601640]
 8. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, Geller BM, Abraham LA, Taplin SH, Dignan M, Cutter G, Ballard-Barbash R. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 2003;138(3):168–75. [PubMed: 12558355]
 9. Yankaskas BC, Taplin SH, Ichikawa L, Geller BM, Rosenberg RD, Carney PA, Kerlikowske K, Ballard-Barbash R, Cutter GR, Barlow WE. Association between mammography timing and measures of screening performance in the United States. *Radiology* 2005;234(2):363–73. [PubMed: 15670994]
 10. American College of Radiology e. American College of Radiology (ACR) Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas). *Am Coll Radiol*. 2003
 11. Liang KY, Zeger SL. Regression analysis for correlated data. *Annu Rev Public Health* 1993;14:43–68. [PubMed: 8323597]
 12. Miglioretti DL, Heagerty PJ. Marginal modeling of multilevel binary data with time-varying covariates. *Biostatistics* 2004;5(3):381–98. [PubMed: 15208201]
 13. Miglioretti DL, Heagerty PJ. Marginal modeling of nonnested multilevel data using standard software. *Am J Epidemiol* 2007;165(4):453–63. [PubMed: 17121864]
 14. Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society Series B, Statistical Methodology* 2006;68(5):859–72.
 15. Egger JR, Cutter GR, Carney PA, Taplin SH, Barlow WE, Hendrick RE, D'Orsi CJ, Fosse JS, Abraham L, Elmore JG. Mammographers' perception of women's breast cancer risk. *Med Decis Making* 2005;25(3):283–9. [PubMed: 15951455]
 16. Rohan TE, Negassa A, Chlebowski RT, Habel L, McTiernan A, Ginsberg M, Wassertheil-Smoller S, Page DL. Conjugated equine estrogen and risk of benign proliferative breast disease: a randomized controlled trial. *J Natl Cancer Inst* 2008;100(8):563–71. [PubMed: 18398105]
 17. Collins LC, Baer HJ, Tamimi RM, Connolly JL, Colditz GA, Schnitt SJ. The influence of family history on breast cancer risk in women with biopsy-confirmed benign breast disease: results from the Nurses' Health Study. *Cancer* 2006;107(6):1240–7. [PubMed: 16902983]
 18. Webb PM, Byrne C, Schnitt SJ, Connolly JL, Jacobs T, Peiro G, Willett W, Colditz GA. Family history of breast cancer, age and benign breast disease. *Int J Cancer* 2002;100(3):375–8. [PubMed: 12115556]
 19. Worsham MJ, Raju U, Lu M, Kapke A, Cheng J, Wolman SR. Multiplicity of benign breast lesions is a risk factor for progression to breast cancer. *Clin Cancer Res* 2007;13(18):5474–9. [PubMed: 17875777]
 20. Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J. The influence of clinical information on the accuracy of diagnostic mammography. *Breast Cancer Res Treat* 2004;85(3):223–8. [PubMed: 15111760]

21. Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr. Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks. *Acad Radiol* 1999;6(1):10–5. [PubMed: 9891147]
22. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, Perrin RL, Poller WR, Shah R, Sumkin JH, Wallace LP, Rockette HE. The “laboratory” effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology* 2008;249(1):47–53. [PubMed: 18682584]
23. Rutter CM, Taplin S. Assessing mammographers' accuracy. A comparison of clinical and test performance. *J Clin Epidemiol* 2000;53(5):443–50. [PubMed: 10812315]
24. Taplin S, Abraham L, Barlow WE, Fenton JJ, Berns EA, Carney PA, Cutter GR, Sickles EA, Carl D, Elmore JG. Mammography facility characteristics associated with interpretive accuracy of screening mammography. *J Natl Cancer Inst* 2008;100(12):876–87. [PubMed: 18544742]

Table 1

Demographic and clinical characteristics by mammogram follow-up recommendation and breast cancer outcomes

	Number of Mammograms	Number of Cancers	Recall Rate for Additional Evaluation (per 100 mammograms)	Biopsy Recommendation Rate (per 100 mammograms)	Cancer Rate w/in one year (per 1,000 mammograms)
	N (%)	N (%)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
TOTAL	638,947 (100)	3024 (100)	10.2	1.4	4.7
Age					
30-39	25,071 (3.9)	42 (1.4)	12.5 (12.1,12.9)	1.3 (1.2, 1.5)	1.7 (1.2, 2.2)
40-49	194,581 (30.5)	539 (17.8)	11.7 (11.6,11.9)	1.3 (1.3, 1.4)	2.8 (2.5, 3.0)
50-59	196,636 (30.8)	889 (29.4)	10.4 (10.2,10.5)	1.4 (1.3, 1.4)	4.5 (4.2, 4.8)
60-69	114,212 (17.9)	724 (23.9)	9.1 (9.0,9.3)	1.4 (1.3, 1.4)	6.3 (5.9, 6.8)
70-79	83,513 (13.1)	619 (20.5)	8.0 (7.8,8.1)	1.4 (1.3, 1.5)	7.4 (6.8, 8.0)
80-89	24,934 (3.9)	211 (7.0)	7.3 (7.0,7.7)	1.3 (1.2, 1.5)	8.5 (7.3, 9.6)
Breast Density [†]					
Almost Entirely Fatty	64,321 (10.1)	163 (5.4)	4.6 (4.5,4.8)	0.7 (0.7, 0.8)	2.5 (2.2, 2.9)
Scattered fibroglandular tissue	271,859 (42.5)	1173 (38.8)	8.9 (8.8,9.1)	1.2 (1.2, 1.2)	4.3 (4.1, 4.6)
Heterogeneously dense	247,727 (38.8)	1403 (46.4)	12.7 (12.6,12.9)	1.6 (1.6, 1.7)	5.7 (5.4, 6.0)
Extremely dense	55,040 (8.6)	285 (9.4)	11.4 (11.2,11.7)	1.5 (1.4, 1.6)	5.2 (4.6, 5.8)
Time since Last Mammogram					
≤ 2 years	534,340 (83.6)	2516 (83.2)	9.4 (9.3,9.5)	1.2 (1.2, 1.3)	4.7 (4.5, 4.9)
3-4 years	42,235 (6.6)	225 (7.4)	11.9 (11.6,12.2)	1.6 (1.5, 1.7)	5.3 (4.6, 6.0)
≥ 5 years or No Previous Mammogram	62,372 (9.8)	283 (9.4)	16.0 (15.7,16.3)	2.1 (2.0, 2.3)	4.5(4.0, 5.1)
Clinical History Risk Factors Studied					
Current Hormone Therapy use?					
No	412,053 (68.7)	1755 (62.3)	8.0 (7.7,8.3)	1.0 (0.9, 1.1)	4.3 (4.1, 4.5)
Yes	187,721 (31.3)	1060 (37.7)	10.2 (10.1,10.3)	1.4 (1.3, 1.4)	5.7 (5.3, 6.0)
<i>Missing</i>	39,173	209	10.7 (10.6,10.9)	1.4 (1.3, 1.5)	5.3 (4.6, 6.1)
Any Family History of breast cancer?					
No	520,150 (83.5)	2302 (77.8)	9.9 (9.4,10.3)	1.2 (1.1, 1.4)	4.4 (4.3, 4.6)

	Number of Mammograms	Number of Cancers	Recall Rate for Additional Evaluation (per 100 mammograms)	Biopsy Recommendation Rate (per 100 mammograms)	Cancer Rate w/in one year (per 1,000 mammograms)
	N (%)	N (%)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Yes	103,116 (16.5)	656 (22.2)	10.2 (10.2,10.3)	1.3 (1.3, 1.3)	6.4 (5.9, 6.9)
Missing	15,681	66	10.0 (9.8,10.2)	1.5 (1.4, 1.6)	4.2 (3.2, 5.2)
Ever had a benign breast biopsy?					
No	498,496 (80.9)	2142 (74.0)	10.2 (9.8,10.6)	1.6 (1.5, 1.8)	4.3 (4.1, 4.5)
Yes	117,569 (19.1)	751 (26.0)	9.8 (9.7,9.9)	1.2 (1.2, 1.3)	6.4 (5.9, 6.8)
Missing	22,882	131	11.9 (11.8,12.1)	1.8 (1.7, 1.9)	5.7 (4.8, 6.7)
Combinations of Current HT use, Family History, and benign biopsy					
All Combinations *					
None	314,497 (49.2)	1182 (39.1)	9.8 (9.7,9.9)	1.2 (1.2, 1.2)	3.8 (3.5, 4.0)
Current HT use only	127,026 (19.9)	629 (20.8)	10.1 (9.9,10.3)	1.2 (1.2, 1.3)	5.0 (4.6, 5.3)
Family History only	60,362 (9.4)	326 (10.8)	9.2 (9.0,9.4)	1.3 (1.2, 1.4)	5.4 (4.8, 6.0)
Benign Biopsy Only	60,371 (9.4)	336 (11.1)	11.6 (11.4,11.9)	1.8 (1.7, 1.9)	5.6 (5.0, 6.2)
Current HT use and Family History	19,493 (3.1)	136 (4.5)	10.5 (10.1,10.9)	1.5 (1.3, 1.7)	7.0 (5.8, 8.2)
Current HT use and Benign Biopsy	33,937 (5.3)	221 (7.3)	12.7 (12.4,13.1)	1.8 (1.6, 1.9)	6.5 (5.7, 7.4)
Family History and Benign Biopsy	15,996 (2.5)	120 (4.0)	11.0 (10.5,11.4)	1.9 (1.7, 2.1)	7.5 (6.2, 8.8)
All 3 risk factors	7,265 (1.1)	74 (2.4)	13.2 (12.4,14.0)	2.2 (1.9, 2.5)	10.2 (7.9, 12.5)
Risk Factor Score (# of Risk Factors) *					
0	314,497 (49.2)	1182 (39.1)	9.8 (9.7,9.9)	1.2 (1.2, 1.2)	3.8 (3.5, 4.0)
1	247,759 (38.8)	1291 (37.7)	10.3 (10.1,10.4)	1.4 (1.3, 1.4)	5.2 (4.9, 5.5)
2	69,426 (10.9)	477 (15.8)	11.7 (11.5,11.9)	1.7 (1.6, 1.8)	6.9 (6.3, 7.5)
3	7,265 (1.1)	74 (2.4)	13.2(12.4, 14.0)	2.2 (1.9, 2.5)	10.2 (7.9, 12.5)

* Risk factor score and all combinations are based on available clinical history at time of mammogram

† Breast Density is a risk factor assigned by radiologist and not part of the clinical history presented to the radiologist

Table 2

Evaluating the relationship between radiologist characteristics and differences between recall rates, sensitivity, specificity, and PPV rates by a woman's level risk factor score

	N	N (%)	RECALL				SENSITIVITY				SPECIFICITY				PPV			
			All		Risk Score		All		Risk Score		All		Risk Score		All		Risk Score	
			Pts	0	1	≥2	Pts	0	1	≥2	Pts	0	1	≥2	Pts	0	1	≥2
TOTAL	134		10.3	9.8	10.3	11.8	81.3	82.3	80.8	80.6	90.1	90.5	90.1	88.6	3.8	3.2	4.1	4.8
Radiologist Demographics																		
Age																		
34-44	43	(32.1)	11.5	11.0	11.6	13.6	83.0	84.6	82.7	80.1	88.9	89.3	88.8	86.9	3.5	3.0	3.9	4.4
45-54	48	(35.8)	10.3	10.4	10.4	11.9	82.2	82.2	83.0	83.1	90.1	90.5	90.0	88.6	3.8	3.2	4.1	4.9
55+	43	(32.1)	8.6	8.3	8.5	10.0	77.1	77.9	77.8	74.3	91.7	92.0	91.8	90.5	4.1	3.4	4.3	5.4
Radiologist Experience in Breast Imaging																		
Years of mammography experience																		
<10	33	(25.0)	10.8	10.3	10.9	12.6	84.2	84.0	84.6	84.0	89.6	90.0	89.5	88.0	3.7	2.9	4.2	4.8
10-19	57	(43.2)	10.6	10.1	10.7	12.5	82.0	83.6	81.6	79.3	89.7	90.2	89.7	88.0	3.8	3.3	4.1	4.7
≥20	42	(31.8)	8.7	8.5	8.6	9.8	76.2	77.2	74.9	77.2	91.6	91.8	91.7	90.7	3.8	3.1	4.0	5.3
<i>Missing</i>	2																	
No. of mammograms interpreted in one year																		
≤1000	36	(27.3)	7.5	7.0	7.6	9.0	75.2	76.0	76.8	70.3	92.8	93.3	92.7	91.4	4.4	3.7	4.7	5.2
1001-2000	50	(37.9)	11.3	10.7	11.5	13.0	81.8	81.6	82.3	80.7	89.1	89.5	88.9	87.5	3.4	2.7	4.0	4.3
>2000	46	(34.8)	10.0	9.7	10.0	11.7	82.0	82.0	80.5	81.1	90.3	90.6	90.6	88.8	3.9	3.4	4.1	5.2
Practice																		
Primary Affiliation with Academic Medical Center																		
No	126	(94.0)	7.0	6.7	7.0	7.9	75.8	75.4	71.2	87.1	93.4	93.5	93.4	92.8	5.2	4.0	5.5	8.9
Yes	8	(6.0)	10.4	9.9	10.4	12.1	81.6	82.7	81.4	79.4	90.0	90.3	89.9	88.4	3.7	3.1	4.1	4.7

* Unadjusted rates are presented throughout the table

Definition:

1. Bold indicates statistical significance ($P < 0.05$) from a GEE model adjusted for patient age, Breast density, and site. Bold in Total row indicates that the risk factor score has a statistically significant outcome compared to both other risk factor scores. A single bold value in all patient columns indicates that the bolded radiologist characteristic value has a statistically significant difference in outcome response compared to all other categories of that radiologist characteristic. Two bold values in all patient columns indicate that the two bolded radiologist characteristic values have a statistically different outcome response from each other, but not to other categories. No bolding within the stratified risk factor by radiologist characteristics indicates no statistically significant interactions.
 2. Recall: positive assessment per 100 mammograms, sensitivity: positive assessment among women with cancer per 100 mammograms, specificity: negative assessment among women without cancer per 100 mammograms, and PPV(positive predictive value): cancer among women with a positive assessment per 100 mammograms.
- Study Population: 134 radiologists, 638,947 screening mammograms, 65,147 total positive assessments, 3024 cancers which 2457 had positive assessments, and 625,923 non-cancers with 573,206 negative assessments.

Table 3

Conditional Logistic regression models examining the association of woman level risk factors within a radiologist on mammography interpretive performance measures

	RECALL		SENSITIVITY		SPECIFICITY		PPV	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Women-level Risk Factors								
Clinical Risk Factor Score								
0	1		1		1		1	
1	1.17	(1.15, 1.19)	1.00	(0.80, 1.24)	0.86	(0.84, 0.88)	1.08	(0.99, 1.19)
≥2	1.43	(1.40, 1.47)	1.07	(0.81, 1.42)	0.70	(0.68, 0.72)	1.12	(0.99, 1.26)
Breast Density								
Almost Entirely Fatty	0.54	(0.52, 0.56)	1.55	(0.86, 2.77)	1.85	(1.77, 1.93)	1.02	(0.84, 1.22)
Scattered fibroglandular tissue	1		1		1		1	
Heterogeneously dense	1.38	(1.36, 1.41)	0.58	(0.45, 0.73)	0.73	(0.71, 0.74)	1.03	(0.94, 1.13)
Extremely dense	1.17	(1.14, 1.21)	0.29	(0.21, 0.41)	0.86	(0.83, 0.89)	1.10	(0.93, 1.30)
Age								
Increase in 10 years of age	0.95	(0.94, 0.96)	1.12	(1.02, 1.22)	1.07	(1.06, 1.08)	1.63	(1.57, 1.69)
Time since Last Mammogram								
≤ 2 years	1		1		1		1	
3 - 4 years	1.32	(1.28, 1.37)	1.83	(1.18, 2.86)	0.76	(0.74, 0.79)	1.22	(1.05, 1.43)
≥ 5 years or No Previous Mammogram	1.84	(1.80, 1.89)	2.27	(1.49, 3.46)	0.55	(0.53, 0.56)	1.00	(0.87, 1.15)

*Results from four multivariable conditional logistic regression models including the Clinical Risk factor score, breast density, linear effect of age, and time since last mammogram on outcomes recall, sensitivity, specificity, and PPV.

Definition:

1. Clinical Risk Factor Score measures the number of clinical risk factors including HT use, family history, and prior biopsy.

2. Bold indicates a statistically significant ($P < 0.05$) odds ratio.

3. OR (odds ratio) estimates the strength of association between having a given covariate-level compared to reference level (e.g. 1 clinical risk factor compared to 0 clinical risk factors) for a particular outcome (recall: positive assessment, sensitivity: positive assessment among women with cancer, specificity: negative assessment among women without cancer, and PPV(positive predictive value): cancer among women with a positive assessment) adjusted for other covariates in the model.

Study Population: 134 radiologists, 638,947 screening mammograms, 65,147 total positive assessments, 3024 cancers which 2457 had positive assessments, and 625,923 non-cancers with 573,206 negative assessments.

Conditional Logistic regression models examining the association of each combination of woman level risk factors within a radiologist on mammography interpretive performance measures

Table 4

Clinical Risk Factors	RECALL		SENSITIVITY		SPECIFICITY		PPV	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
No Risk Factors	1		1		1		1	
1 Risk Factor								
Current HT use	1.20	(1.18, 1.23)	1.08	(0.82, 1.43)	0.83	(0.81, 0.85)	0.99	(0.88, 1.10)
Family History	0.99	(0.96, 1.02)	0.84	(1.60, 1.16)	1.02	(0.99, 1.05)	1.37	(1.19, 1.58)
Benign Biopsy	1.29	(1.25, 1.33)	1.03	(0.73, 1.44)	0.78	(0.75, 0.80)	1.05	(0.91, 1.21)
2 Risk Factors								
Current HT use and Family History	1.29	(1.23, 1.35)	0.85	(0.54, 1.35)	0.78	(0.74, 0.82)	1.18	(0.95, 1.46)
Current HT use and Benign Biopsy	1.58	(1.52, 1.64)	1.15	(0.77, 1.70)	0.63	(0.61, 0.65)	0.92	(0.78, 1.09)
Family History and Benign Biopsy	1.25	(1.19, 1.32)	0.96	(0.58, 1.58)	0.81	(0.77, 0.85)	1.36	(1.09, 1.70)
All 3 risk factors	0.64	(0.58, 0.72)	1.96	(0.67, 5.74)	1.56	(1.40, 1.74)	0.99	(0.85, 1.14)

*Results from four multivariable conditional logistic regression models including indicators for all clinical risk factor combinations, breast density, linear effect of age, and time since last mammogram on outcomes recall, sensitivity, specificity, and PPV.

Definition:

1. Bold indicates a statistically significant ($P < 0.05$) odds ratio.
2. OR (odds ratio) estimates the strength of association between having a given covariate-level compared to reference level (e.g. 1 clinical risk factor compared to 0 clinical risk factors) for a particular outcome (recall: positive assessment, sensitivity: positive assessment among women with cancer, specificity: negative assessment among women without cancer, and PPV(positive predictive value): cancer among women with a positive assessment) adjusted for other covariates in the model.

Study Population: 134 radiologists, 638,947 screening mammograms, 65,147 total positive assessments, 3024 cancers which 2457 had positive assessments, and 625,923 non-cancers with 573,206 negative assessments.

Table 5

Expected Positive Predictive Value (PPV) based on observed cancer rate and assuming a fixed sensitivity and specificity across clinical risk factor score.

	Cancer Rate w/in One Year of Screening (per 1,000 mammograms)	PPV Expected Odds Ratio ¹	PPV Observed Unadjusted Odds Ratio (95% CI)	PPV Observed Adjusted ² Odds Ratio (95% CI)
Clinical Risk Factor Score				
0	3.8	1	1	1
1	5.2	1.39	1.31 (1.19, 1.44)	1.08 (0.99, 1.19)
2+	7.2	1.92	1.55 (1.39, 1.73)	1.12 (0.99, 1.26)

¹ Expected PPV Odds Ratio comparing Group B to Group A is calculated as

$$\frac{CancerRate_B / (1 - CancerRate_B)}{CancerRate_A / (1 - CancerRate_A)}$$

² Adjusted Odds Ratios from the multivariable model including the Clinical Risk factor score, breast density, linear effect of age, and time since last mammogram.