Title

Prioritizing Screening Mammograms for Immediate Interpretation and Diagnostic Evaluation on the Basis of Risk for Recall

Journal

Journal of the American College of Radiology, 20(3)

ISSN

1546-1440

Authors

Ho, Thao-Quyen H
Bissell, Michael CS
Lee, Christoph I
et al.

Publication Date

2023-03-01

DOI

10.1016/j.jacr.2022.09.030

Peer reviewed

HHS Public Access

Author manuscript

J Am Coll Radiol. Author manuscript; available in PMC 2023 March 28.

Published in final edited form as:

J Am Coll Radiol. 2023 March ; 20(3): 299–310. doi:10.1016/j.jacr.2022.09.030.

Prioritizing Screening Mammograms for Immediate Interpretation and Diagnostic Evaluation on the Basis of Risk for Recall

Thao-Quyen H. Ho, MD, PhD[a,b,c], **Michael C. S. Bissell, PhD**[a], **Christoph I. Lee, MD, MS**[d,e,f,h,i], **Janie M. Lee, MD**[d,f,g], **Brian L. Sprague, PhD**[j], **Anna N. A. Tosteson, ScD**[k], **Karen J. Wernli, PhD**[l,m], **Louise M. Henderson, PhD**[n,o], **Karla Kerlikowske, MD**[p,q,r], **Diana L. Miglioretti, PhD**[a,m,s]

[a]Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, School of Medicine, Davis, California.

[b]Breast Imaging Unit, Diagnostic Imaging Center, Tam Anh General Hospital, Ho Chi Minh City, Vietnam.

[c]Department of Training and Scientific Research, University Medical Center, Ho Chi Minh City, Vietnam.

[d]Breast Imaging, Department of Radiology, University of Washington School of Medicine, Seattle, Washington.

[e]Department of Health Systems and Population Health, University of Washington School of Public Health, Seattle, Washington.

[f]Hutchinson Institute for Cancer Outcomes Research, Seattle, Washington.

[g]Breast Imaging, Fred Hutchinson Cancer Center, Seattle, Washington.

[h]Northwest Screening and Cancer Outcomes Research Enterprise, University of Washington, Seattle, Washington.

[i]Deputy Editor, *JACR*.

[j]Department of Surgery, Office of Health Promotion Research, Larner College of Medicine at the University of Vermont and Co-Leader, Cancer Control and Population Health Sciences Program, University of Vermont Cancer Center, Burlington, Vermont.

[k]The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth and Associate Director for Population Sciences, Dartmouth Cancer Center, Lebanon, New Hampshire.

[l]Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, California.

Corresponding author and reprints: Diana L. Miglioretti, PhD, Department of Public Health Sciences, UC Davis School of Medicine, One Shields Avenue, Med Sci 1C, Davis, CA 95616; dmiglioretti@ucdavis.edu.
Drs Ho and Bissell contributed equally to this work.

ADDITIONAL RESOURCES

Additional resources can be found online at: https://doi.org/10.1016/j.jacr.2022.09.030

[m]Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, Washington.

[n]Department of Radiology, University of North Carolina, Chapel Hill, North Carolina.

[o]Cancer Epidemiology Program, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina.

[p]Departments of Medicine and Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California.

[q]General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco, San Francisco, California.

[r]Women's Health Comprehensive Clinic, and Director, Advanced Post-doctoral Fellowship in Women's Health, San Francisco Veterans Affairs Health Care System, San Francisco, California.

[s]Biostatistics and Population Sciences and Health Disparities Program, University of California, Davis, Comprehensive Cancer Center, Davis, California.

Abstract

Purpose: The aim of this study was to develop a prioritization strategy for scheduling immediate screening mammographic interpretation and possible diagnostic evaluation.

Methods: A population-based cohort with screening mammograms performed from 2012 to 2020 at 126 radiology facilities from 7 Breast Cancer Surveillance Consortium registries was identified. Classification trees identified combinations of clinical history (age, BI-RADS® density, time since prior mammogram, history of false-positive recall or biopsy result), screening modality (digital mammography, digital breast tomosynthesis), and facility characteristics (profit status, location, screening volume, practice type, academic affiliation) that grouped screening mammograms by recall rate, with ≥12/100 considered high and ≥16/100 very high. An efficiency ratio was estimated as the percentage of recalls divided by the percentage of mammograms.

Results: The study cohort included 2,674,051 screening mammograms in 925,777 women, with 235,569 recalls. The most important predictor of recall was time since prior mammogram, followed by age, history of false-positive recall, breast density, history of benign biopsy, and screening modality. Recall rates were very high for baseline mammograms (21.3/100; 95% confidence interval, 19.7–23.0) and high for women with ≥5 years since prior mammogram (15.1/100; 95% confidence interval, 14.3–16.1). The 9.2% of mammograms in subgroups with very high and high recall rates accounted for 19.2% of recalls, an efficiency ratio of 2.1 compared with a random approach. Adding women <50 years of age with dense breasts accounted for 20.3% of mammograms and 33.9% of recalls (efficiency ratio = 1.7). Results including facility-level characteristics were similar.

Conclusions: Prioritizing women with baseline mammograms or ≥5 years since prior mammogram for immediate interpretation and possible diagnostic evaluation could considerably reduce the number of women needing to return for diagnostic imaging at another visit.

Keywords

Screening mammography; recall rate; immediate interpretation; Breast Cancer Surveillance Consortium

INTRODUCTION

Screening mammography reduces breast cancer mortality, but this benefit must be weighed against potential harms [1,2]. The most common potential harm for women is being recalled for additional workup [2,3], which occurs in approximately 12% of screening mammograms, the majority of which are false-positive findings [4]. Overall, 36% to 56% of women may expect to experience at least one false-positive recall over 10 years of annual or biennial screening [5]. Being recalled for additional imaging leads to anxiety, requires women to return for another visit, and incurs financial strain and opportunity costs [2,3,6,7]. Having to schedule additional imaging on a future date also contributes to delays in diagnosis for women with cancer and the possibility of women not returning that may result in even longer delays in diagnosis [6].

Some facilities now offer immediate interpretation of screening mammograms while women wait and diagnostic imaging workup as needed during the same visit. This approach may greatly reduce psychological anxiety associated with the anticipation of additional imaging evaluation and decrease gaps in follow-up care among women who would not return for additional imaging at a later date [6,8]. For facilities that can only offer immediate interpretation to a subset of women, prioritizing those most likely to be recalled may considerably reduce the number of women being recalled for diagnostic imaging at a follow-up visit. Prior research has shown higher recall rates on baseline mammograms and in younger women or those with longer screening intervals, family history of breast cancer, heterogeneously dense breasts, or history of benign biopsy [9–11]. However, no previous study has quantified which combinations of risk factors predict those at highest risk for a recall.

Our goal was to identify subgroups of women at the highest risk for being recalled for additional workup of a screening mammogram. Using data on a large cohort from seven Breast Cancer Surveillance Consortium (BCSC) registries, we created a hierarchy of subgroups according to recall rate that could aid facilities at the time of scheduling in prioritizing women for immediate screening interpretation and potential same-day diagnostic workup according to capacity.

METHODS

Study Setting and Cohort

This report follows Strengthening the Reporting of Observational Studies in Epidemiology guidelines for cohort studies. Prospective data were collected at seven US-based BCSC registries (www.bcsc-research.org): Carolina Mammography Registry, Kaiser Permanente Washington, Metro Chicago Breast Cancer Registry, New Hampshire Mammography Network, Sacramento Area Breast Imaging Registry, San Francisco Mammography

Registry, and Vermont Breast Cancer Surveillance System. Registries collect individual-level characteristics and clinical information from community radiology facilities [4] and link to state or regional cancer registries and pathology databases for complete capture of breast cancer diagnoses and benign breast biopsy results. We included all digital mammographic and digital breast tomosynthesis screening examinations performed from 2012 to 2020 among women aged 18 to 89 years without personal histories of breast cancer [4] and no missing data on the covariates of interest. The final study population included 2,674,051 screening mammograms from 925,777 women interpreted by 644 radiologists across 126 breast imaging facilities.

BCSC registries and the statistical coordinating center received institutional review board approval for active or passive consenting processes or a waiver of the requirement to obtain consent to enroll participants, link, and pool data, and perform analysis. Procedures were HIPAA compliant, and a federal certificate of confidentiality protects the identities of participants.

Characteristics of Mammograms in the Cohort and Facilities

Age, time since prior mammogram (if any), first-degree family history of breast cancer, and history of breast biopsy were collected from self-administered health history questionnaires at the time of mammography or extracted from electronic health records. Time since prior mammogram was categorized as first mammogram, 1 year (9–18 months), 2 years (19–30 months), 3 years (31–42 months), 4 years (43–54 months), or 5 years or longer (>54 months). Mammograms were considered first mammograms if no previous mammogram was identified in the database and individuals self-reported that they did not have prior mammograms. Mammograms were considered subsequent examinations if a previous mammogram was identified in the database or a prior mammogram date was self-reported. Mammograms with no previous mammogram in the database and missing self-reported information on time since last mammogram were excluded. Examinations with mammograms in the prior 9 months were excluded to avoid mis-classifying diagnostic mammograms as screening. History of benign breast biopsy result and high-risk lesion diagnoses (atypical hyperplasia or lobular carcinoma in situ) were obtained via self-report and linkage with pathology databases. Diagnoses of lobular carcinoma in situ were also obtained by linkage with state or regional cancer registries, regardless of enrollment in the BCSC at the time of diagnosis. Self-reported race/ethnicity was categorized as Hispanic/Latina or non-Hispanic Asian, Black, White, other, or multiracial (if non-Hispanic with more than one self-reported race).

Radiologists reported assessments according to ACR BI-RADS® terminology and BI-RADS breast density as part of the clinical interpretation as almost entirely fatty, scattered fibroglandular densities, heterogeneously dense, or extremely dense [12]. Breast density from the most recent prior mammogram was used, given that would be available at the time of scheduling, and if not available, breast density at the time of the screening examination was used. History of false-positive recall was defined as a screening mammogram with a positive initial assessment within 5 years before the screening examination.

We collected the following facility-level characteristics: profit status, location, screening volume, radiology practice type, and academic affiliation. Screening volume was computed as median annual screening volume in our study and categorized into approximate quartiles as <1,000, 1,000 to 2,999, 3,000 to 4,999, and ≥5,000.

Mammographic Outcomes

The primary outcome was recall, defined as a positive BI-RADS initial assessment on the screening examination of 0 (needs additional imaging evaluation), 3 (probably benign finding), 4 (suspicious abnormality), or 5 (highly suggestive of malignancy). Screening examinations with same-day diagnostic imaging performed (0.3% of screens) were considered recalls. Recall rate was calculated as the number of positive initial assessments divided by the total number of screening mammograms.

Statistical Analysis

We summarized characteristics of the mammograms included and those recalled. We estimated recall rates for clinical factors and estimated 95% confidence intervals (CIs) using generalized estimating equations with a working independence correlation structure to account for correlation among mammograms within the same facility [13]. To estimate the relative efficiency of prioritizing mammograms using a risk-based approach compared with a random, non-risk-based approach, an efficiency ratio was estimated as the percentage of recalls divided by the percentage of mammograms within each subgroup, with a ratio of 1.0 for the overall study population.

Classification trees [14] were used to identify subgroups with similar recall rates. For the primary analysis, we fit separate models for first screening mammograms and mammograms with ≥5 years since prior mammogram (model 1) and subsequent screening mammograms with 1, 2, 3, or 4 years since prior mammogram for added flexibility in the classification tree algorithm (model 2). Both models included time since prior mammogram, age group (<40, 40–49, 50–59, 60–69, and ≥70 years), first-degree family history of breast cancer, screening modality, history of benign biopsy result, and history of high-risk breast lesion. History of false-positive recall and breast density were included in model 2 only, given that these variables are not applicable for baseline mammograms and not meaningful for women with ≥5 years since prior mammogram. We performed a secondary analysis adding facility-level covariates to both models.

We set a maximum tree depth (ie, number of levels in the decision tree) of 10, a maximum number of leaves (ie, final subgroups) of 10, and a minimum leaf size of at least 0.5% of the overall cohort. To internally validate each model, we split the study populations into two random samples, fit the classification tree models to each sample, and found no clinically meaningful differences with the overall model. Random forests of 10,000 trees for each model ranked variable importance similar to the single classification tree models [15,16].

We combined the subgroups (leaves) into four risk groups on the basis of their recall rates: very high (≥16), high (12 to <16), moderate (8 to <12), and low (<8) (per 100 mammograms). Thresholds and logical grouping of similar subgroups were chosen on the basis of clinical guidelines adopted by the ACR (for considering >12% high) [12,17],

observed recall rates, and clinical expertise. We plotted proportion of cases followed curves [18], that is, the cumulative percentage of recalls by the cumulative percentage of mammograms when subgroups are ordered from highest to lowest recall rates, both overall and for each facility that contributed at least 10,000 screening mammograms to ensure sufficient sample size for estimating recall rate in the smallest subgroups. The 45° reference line represents the cumulative percentage of recalls expected using a random, non-risk-based approach.

We conducted several sensitivity analyses: (1) restricting the analysis to the most recent 5 years of data to evaluate the effects of potential changes in recall rates over time and separately including (2) BCSC 5-year risk score [19] and (3) race/ethnicity in the models to determine if either improved prediction of recall beyond other risk factors. In all three cases, results were similar to the main analysis.

Statistical analyses were performed using SAS version 9.4 (SAS Institute), R version 4.0.2 (R Foundation for Statistical Computing), and RStudio version 1.3.1056 (Posit).

RESULTS

More than three-fourths of mammograms were in women ≥50 years of age (78.9%), 69.1% of women identified as White, 18.5% had first-degree family histories of breast cancer, 16.7% had prior false-positive recalls, 21.3% had prior benign biopsy results, and 0.5% had histories of high-risk lesion (Table 1). Nearly half of mammograms were in women with dense breasts, two-thirds were annual screening mammograms, and 32.1% were digital breast tomosynthesis examinations. The overall recall rate was 8.8 per 100 mammograms (95% CI, 8.3–9.4). Recall rates were higher at younger ages; among women with histories of false-positive recall, prior benign biopsy results, or high-risk lesions; among women with dense breasts; for baseline mammograms and longer time since previous mammogram; and for digital mammograms. The unadjusted recall rates of Black, Hispanic, and other or multiracial women ranged from 9.2 to 9.6 per 100, higher than among White women (8.8/100; 95% CI, 8.2–9.4), whereas Asian women had the lowest recall rate (7.8/100; 95% CI, 6.9–8.9). Recall rates were similar for those with (8.8/100; 95% CI, 8.3–9.4) and without (8.8/100; 95% CI, 8.2–9.4) first-degree family histories of breast cancer. The largest efficiency ratios were among baseline mammograms (2.4), women <50 years of age (1.4–1.9), and women with ≥5 years since prior mammogram (1.7).

Most facilities were not for profit (65.4%), located within hospitals (50.5%), full diagnostic radiology practices (64.5%), and nonacademic (86.9%) (Table 2). Most screening mammograms were from facilities with median annual screening volumes ≥5,000 (61.5%).

Table 3 summarizes the recall risk groups, ordered from highest to lowest recall rate, in which some subgroups have been combined using clinical expertise to simplify scheduling rules at the facility level. Expanded subgroups details are provided in Supplementary Table 1. The classification tree algorithm ordered the predictors from highest to lowest variable importance for model 1 as baseline versus ≥5 years since prior mammogram, age group, and screening modality and for model 2 as age group, history of false-positive recall,

breast density, time since prior mammogram, history of benign biopsy result, and screening modality. Neither model selected first-degree family history of breast cancer or history of high-risk lesion.

Recall rates were considered very high for baseline mammograms (21.3/100; 95% CI, 19.6–23.2) and high for women with ≥5 years since prior mammogram (15.1/100; 95% CI, 14.3–16.1) (Table 2). Recall risk was considered moderate for women <50 years of age and those with 1 to 4 years since their prior mammogram with either dense breasts (11.7/100; 95% CI, 11.0–12.5) or scattered fibroglandular densities (9.7/100; 95% CI, 8.9–10.5). Women aged 50 or older with 1 to 4 years since prior mammogram and histories of false-positive recall were also at moderate risk for recall (9.5/100; 95% CI, 8.9–10.1). Recall rates were considered low otherwise.

The 4.9% of mammograms classified as very high risk for recall accounted for 11.8% of all recalls (efficiency ratio = 2.4) (Fig. 1). Adding the 4.3% of mammograms at high risk for recall, for a total of 9.2% of mammograms, accounted for 19.2% of recalls (cumulative efficiency ratio = 2.1). Adding the 11.1% of mammograms in women <50 years of age with dense breasts who were just below the threshold of high risk, for a total of 20.3% of mammograms, accounted for 33.9% of recalls (cumulative efficiency ratio = 1.7). Adding the remaining 18.0% of mammograms at moderate risk for recall in women <50 years of age with 1 to 4 years since prior mammogram and scattered fibroglandular densities, and women ≥50 years of age with 1 to 4 years since prior mammogram and histories of false-positive recall, for a total of 38.3% of mammograms, accounted for 53.4% of recalls. The 61.7% of mammograms with low risk for recall accounted for the remaining 46.6% of recalls. The percentage of mammograms considered very high or high varied from 3.2% to 21.8% across large facilities, with cumulative efficiency ratios varying from 1.3 to 3.2.

In our secondary analysis, no facility-level characteristics were selected for the model for subsequent screens. The model for first screens or screens with ≥5 years since prior mammography selected profit status, facility location, screening volume, and practice type; however, recall rates for all resulting subgroups that split on a facility-level characteristic remained greater than the 16% very high risk threshold (Supplementary Table 2); thus, the final groups, when combined across similar covariate combinations and recall risk, were similar to the main analysis (Supplemental Table 3). Unlike the primary analysis, women <50 years of age were not further subdivided by breast density.

DISCUSSION

We used data from 126 US-based radiology facilities in the BCSC to identify subgroups of mammograms with combinations of individual characteristics (age, family history of breast cancer, prior false-positive recall, prior benign biopsy result, and breast density), time since prior mammogram, and screening modality associated with very high, high, moderate, and low recall rates. Radiology facilities could use our results to schedule women for immediate mammographic interpretation and possible same-day diagnostic evaluation on the basis of their capacity to do so. For example, facilities with the capacity to offer immediate interpretation to only 5% of mammograms could prioritize first mammograms, which

consistently had the highest recall rates of 20% to 23%. Facilities with the capacity to offer immediate interpretation to 9% of mammograms could add mammograms in women with ≥5 years since prior mammogram. Those with the capacity to offer immediate interpretation to 20% of screening mammograms could add women <50 years of age with 1 to 4 years since prior mammogram and dense breasts. If facilities prefer a simpler prioritization method to alleviate the burden on the scheduling team eliciting health history questions from the patient or the medical records, they could additionally prioritize all women <50 years of age regardless of breast density, including the small group of women <50 years of age with fatty breasts (0.8%) despite being at low risk for recall, thereby considerably reducing the number of women being recalled for diagnostic imaging at a follow-up visit (40.2%) by performing immediate interpretation of only 26% of mammograms.

Although some facility-level characteristics were associated with risk for recall among first screens and screens with ≥5 years since prior mammography, recall rates for the resulting subgroups remained greater than the 16% very high risk threshold. Thus, our prioritization approach may be applied regardless of the facility's profit status, location, screening volume, practice type, and academic affiliation. Unlike the primary analysis, women <50 years of age were not further subdivided by breast density, which supports our recommendation that facilities with sufficient capacity may consider offering immediate interpretation to all women <50 years of age.

Immediate interpretation of screening mammograms offers several benefits compared with batch reading. Previous research suggests that women recalled for additional imaging may experience elevated anxiety and distress, even if transient [3,20]. In a controlled trial, women with abnormal mammographic findings experienced higher levels of anxiety as much as 3 months later compared with those with normal findings, but those whose mammograms were interpretated immediately and received same-day diagnostic workup reported less anxiety [21]. The majority of recalls will result in false-positive findings [4], which some studies showed could lead to women's being less likely to reattend screening [22], although other studies revealed no difference in rescreening rates [23], and others showed that women with prior false-positive findings were more likely to return [24]. For women who may be discouraged from returning for future screening by a false-positive result, this harm may be partially mitigated by immediate screening interpretation and same-day diagnostic evaluation. Other benefits of immediate read protocols are the ability to repeat imaging during the same appointment when image quality is poor [25] and potential improvements in radiologist performance due to immediate feedback of diagnostic workup from recalled screening mammograms [26].

A potential health equity–related benefit of immediate interpretation is an increase in the percentage of women with timely receipt of additional imaging after abnormal screening results [25]. Black, Hispanic, and Asian women are more likely to experience diagnostic delays between an abnormal screening mammogram and definitive tissue diagnosis, nearly double the rate in some cases, compared with White women [27–30]. Delays may be due to barriers in scheduling including lack of translators, out-of-pocket costs from copays or lack of health insurance, the need to take additional time off work, lack of transportation or childcare, or structural racism factors from chronic underinvestment in neighborhoods

and facilities serving racial and ethnic minority groups [28]. Immediate interpretation and same-day diagnostic imaging offers a solution to barriers associated with having to return for another visit to work up an abnormal screening examination; however, unless facilities can also offer same-day breast biopsy for the 17% of additional evaluations that result in biopsy recommendations [31], some women may still experience delays. Two studies at a single institution that serve a diverse population showed that racial and ethnic disparities in time to diagnostic imaging after abnormal screening mammographic findings improved with immediate interpretation [25] and no evidence of racial and ethnic disparities in time from biopsy recommendation to biopsy being performed after implementation of a same day biopsy program [32]. However, if facilities with sufficient resources are more likely to offer immediate interpretation than those serving marginalized communities, disparities could widen. Unadjusted recall rates were higher among Black, Hispanic, and other or multiracial women compared with White women and lower among Asian women; however, results were similar when adding race/ethnicity to the models. This suggests that racial or ethnic differences in recall rates may be accounted for by the other covariates we investigated and lends weight to the importance of prioritizing immediate interpretation and diagnostic evaluation of screening mammograms at high risk for recall to improve health equity [25].

Although immediate interpretation of screening mammograms alleviates the need for a woman to return on a different day for diagnostic imaging [25], it also introduces challenges for implementation. Revising a facility's clinical template to support immediate interpretation of screening mammograms likely involves increasing the length of screening mammography appointments to include additional time for imaging interpretation and results communication. Without extending clinic hours and staffing, increasing screening appointment lengths may reduce a clinic's screening mammography capacity. For facilities only offering screening mammography without an on-site radiologist, radiology IT and workflow updates may be needed to enable off-site immediate interpretation and partnership with another facility with same-day access for diagnostic workup. For facilities offering screening and diagnostic imaging services, additional clinical template revisions may be needed to support same day add-on diagnostic mammography and ultrasound appointments. Accommodating same day add-on image-guided biopsies would warrant further revisions to a clinic's interventional appointment template. Although these clinical template revisions support timely resolution of screening episodes, holding these appointments for same-day use may result in decreased overall clinic volume if the same-day appointments are not fully used.

In addition to volume, capacity, access, and efficiency considerations, some studies showed that radiologists are more likely to request additional views if the woman is waiting compared with batch reading [33,34], whereas another study showed no difference [35]. There may be cost considerations associated with immediate interpretation as well. Older studies estimated marginal costs of immediate interpretation to be as low as $4 per examination for facilities that do not require additional staffing or space and as high as $28 per examination for facilities operating at full capacity and requiring additional space [36,37]. Two studies from a single academic medical center successfully changed workflow without increasing staffing or space during a period of coronavirus disease 2019 (COVID-19)–induced service reduction [25,32]; however, maintaining this workflow with

a resumption of pre-COVID-19 service volumes may be more challenging. Costs may also be higher now, with cost-of-service provisioning and number of staff members increasing for a comparable service volume. Any additional operating costs associated with adopting immediate interpretation during the ongoing COVID-19 pandemic could potentially exacerbate disparities in access if only facilities serving more affluent communities can afford these additional service costs or if costs are passed directly on to the woman. From the patient's perspective, studies in nonmammography settings show a preference for immediate communication of results even if it means waiting longer [25,32]. Among patients who preferred immediate interpretation, 78% were willing to wait 30 to 60 min for results; however, only 11% of patients were willing to pay an additional $25 in direct costs, and 60% were not willing to pay any additional costs [36].

Our study is one of the largest evaluations of recall rates using data representative of recent US practice. Nonetheless, our study had several limitations. We did not include women with personal histories of breast cancer because diagnostic views were often performed during the screening examination, and we could not determine if these diagnostic views were part of surveillance imaging or for the evaluation of a new finding. Although prior research has shown higher recall rates when comparison images are unavailable to the reading radiologist [38], most facilities in the BCSC do not provide data on comparison image availability. However, among subsequent screening examinations with complete information on comparison images, only 0.8% indicated no comparison films. Other approaches to identify mammograms at high risk for recall should be explored. For example, artificial intelligence algorithms applied directly to the mammographic images while a woman waits may help in determining which screening mammograms are likely to be recalled and thus who should receive immediate interpretation and diagnostic workup [39–41]. We did not evaluate cancer detection rates or which recalled examinations are likely to have true-positive findings. Facilities with a limited capacity to offer immediate interpretation could further prioritize by risk for having a cancer detected [42]. Our models focused on identifying women at high risk for recall and did not consider the risk of not returning if recalled. Future studies should consider how to balance the risk for recall and the risk of not returning when prioritizing immediate interpretation and diagnostic workup. Although we hypothesize our approach for prioritizing immediate interpretation and diagnostic imaging will improve compliance with additional workup on the basis of prior studies [25], we did not directly evaluate this.

In conclusion, time since prior mammogram and age were the most important risk factors associated with high and very high risk for recall, regardless of facility-level characteristics. Scheduling baseline mammograms and women with ≥5 years since prior mammogram for immediate interpretation and diagnostic evaluation could avert one-fifth of recalls for diagnostic imaging at follow-up visits, reducing diagnostic delays and patient anxiety.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Gloria Jackson-Nefertiti, BA, Dianne Johnson, BA, and Jill Budesky, AA, for their contributions as paid patient partners on our funded award. We thank the participating individuals, mammography facilities, and radiologists for the data they have provided for this study. A list of BCSC investigators and staff members can be found at https://www.bcsc-research.org/about/personnel. Information about the BCSC may be found at http://www.bcsc-research.org.

This work was supported through a Patient-Centered Outcomes Research Institute Program Award (PCS-1504-30370). Data collection for this research was additionally supported by the Breast Cancer Surveillance Consortium with funding from the National Cancer Institute (P01CA154292, U54CA163303), the Agency for Healthcare Research and Quality (R01 HS018366-01A1), and the University of Vermont Cancer Center with funds generously awarded by the Lake Champlain Cancer Research Organization (grant 032800). The collection of Sacramento Area Breast Imaging Registry data was supported by the University of California, Davis, Comprehensive Cancer Center, the Placer County Breast Cancer Foundation, and the University of California, Davis, Clinical and Translational Science Center. Cancer and vital status data collection was supported by several state public health departments and cancer registries (http://www.bcsc-research.org/work/acknowledgement.html). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute or its board of governors or methodology committee, the National Cancer Institute, or the National Institutes of Health.

Diana L. Miglioretti declares PCORI funded the research and PCORI, NCI funded the data collection, and royalties from Elsevier. Thao-Quyen H. Ho declares funding from PCORI and NCI. Michael C. S. Bissell declares funding from PCORI and NCI. Christopher I. Lee declares funding from National Cancer Institute; textbook royalties from Oxford University Press, McGraw Hill, Inc., and UpToDate, Inc.; participation with GRAIL, Inc; and personal fees from the American College of Radiology for JACR Deputy Editor duties. Janie M. Lee declares funding from National Cancer Institute and a research grant from GE Healthcare for a separate research study. Brian L. Sprague declares funding from National Cancer Institute; payments to institution from PCORI and Lake Champlain Cancer Research Organization; and grant from the American College of Radiology as payment to institution. Anna N. A. Tosteson declares funding from PCORI to institution for project and NIH/NCI funding to institution for data collection. Karen J. Wernli declares funding from Patient Centered Outcomes Research Institute (PCS-1504-30370) and National Cancer Institute (P01CA154292) and grants from National Cancer Institute and Centers for Disease Control and Prevention. Louise M. Henderson declares funding from NIH and PCORI grant funding to institution. Karla Kerlikowske declares funding from NCI and PCORI funding to institution. The authors are non-partner/non-partnership track/employees.

REFERENCES

1. Mack DS, Lapane KL. Screening mammography among older women: a review of United States guidelines and potential harms. J Womens Health (Larchmt) 2019;28:820–6. [PubMed: 30625008]
2. Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of breast cancer screening: systematic review to update the 2009 U.S. Preventive Services Task Force recommendation. Ann Intern Med 2016;164:256–67. [PubMed: 26756737]
3. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology 2005;14:917–38. [PubMed: 15786514]
4. Lehman CD, Arao RF, Sprague BL, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology 2017;283:49–58. [PubMed: 27918707]
5. Ho TH, Bissell MCS, Kerlikowske K, et al. Cumulative probability of false-positive results after 10 years of screening with digital breast tomosynthesis vs digital mammography. JAMA Netw Open 2022;5: e222440. [PubMed: 35333365]
6. Brocken P, Prins JB, Dekhuijzen PN, van der Heijden HF. The faster the better?—A systematic review on distress in the diagnostic phase of suspected cancer, and the influence of rapid diagnostic pathways. Psychooncology 2012;21:1–10.
7. Harris RP, Sheridan SL, Lewis CL, et al. The harms of screening: a proposed taxonomy and application to lung cancer screening. JAMA Intern Med 2014;174:281–5. [PubMed: 24322781]

8. Barentsz MW, Wessels H, van Diest PJ, et al. Same-day diagnosis based on histology for women suspected of breast cancer: high diagnostic accuracy and favorable impact on the patient. PLoS ONE 2014;9:e103105. [PubMed: 25047134]
9. Lowry KP, Coley RY, Miglioretti DL, et al. Screening performance of digital breast tomosynthesis vs digital mammography in community practice by patient age, screening round, and breast density. JAMA Netw Open 2020;3:e2011792. [PubMed: 32721031]
10. Yankaskas BC, Taplin SH, Ichikawa L, et al. Association between mammography timing and measures of screening performance in the United States. Radiology 2005;234:363–73. [PubMed: 15670994]
11. Nelson HD, O'Meara ES, Kerlikowske K, Balch S, Miglioretti D. Factors associated with rates of false-positive and false-negative results from digital mammography screening: an analysis of registry data. Ann Intern Med 2016;164:226–35. [PubMed: 26756902]
12. D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS® atlas. Reston, Virginia: American College of Radiology; 2013.
13. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.
14. Breiman L, Friedman GH, Stone CJ, Olshen RA. Classification and regression trees. New York: Chapman & Hall; 1993.
15. Dong W, Bensken WP, Kim U, Rose J, Berger NA, Koroukian SM. Phenotype discovery and geographic disparities of late-stage breast cancer diagnosis across U.S. counties: a machine learning approach. Cancer Epidemiol Biomarkers Prev 2022;31:66–76. [PubMed: 34697059]
16. Breiman L. Random forests. Machine Learning 2001;45:5–32.
17. Carney PA, Sickles EA, Monsees BS, et al. Identifying minimally acceptable interpretive performance criteria for screening mammography. Radiology 2010;255:354–61. [PubMed: 20413750]
18. Pfeiffer RM. Extensions of criteria for evaluating risk prediction models for public health applications. Biostatistics 2013;14:366–81. [PubMed: 23087412]
19. Tice JA, Miglioretti DL, Li CS, Vachon CM, Gard CC, Kerlikowske K. Breast density and benign breast disease: risk assessment to identify women at high risk of breast cancer. J Clin Oncol 2015;33:3137–43. [PubMed: 26282663]
20. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. Ann Fam Med 2013;11:106–15. [PubMed: 23508596]
21. Barton MB, Morley DS, Moore S, et al. Decreasing women's anxieties after abnormal mammograms: a controlled trial. J Natl Cancer Inst 2004;96:529–38. [PubMed: 15069115]
22. Dabbous FM, Dolecek TA, Berbaum ML, et al. Impact of a false-positive screening mammogram on subsequent screening behavior and stage at breast cancer diagnosis. Cancer Epidemiol Biomarkers Prev 2017;26:397–403. [PubMed: 28183826]
23. Pinckney RG, Geller BM, Burman M, Littenberg B. Effect of false-positive mammograms on return for subsequent screening mammography. Am J Med 2003;114:120–5. [PubMed: 12586231]
24. Salz T, Richman AR, Brewer NT. Meta-analyses of the effect of false-positive mammograms on generic and specific psychosocial outcomes. Psychooncology 2010;19:1026–34. [PubMed: 20882572]
25. Dontchos BN, Achibiri J, Mercaldo SF, et al. Disparities in same-day diagnostic imaging in breast cancer screening: impact of an immediate-read screening mammography program implemented during the COVID-19 pandemic. AJR Am J Roentgenol 2022;218: 270–8. [PubMed: 34494449]
26. Buist DS, Anderson ML, Smith RA, et al. Effect of radiologists' diagnostic work-up volume on interpretive performance. Radiology 2014;273:351–64. [PubMed: 24960110]
27. Molina Y, Silva A, Rauscher GH. Racial/ethnic disparities in time to a breast cancer diagnosis: the mediating effects of health care facility factors. Med Care 2015;53:872–8. [PubMed: 26366519]
28. Lawson MB, Bissell MCS, Miglioretti DL. Multilevel factors associated with time to biopsy after abnormal screening mammography results by race and ethnicity. JAMA Oncol 2022;8: 1115–26. [PubMed: 35737381]

29. Nguyen KH, Pasick RJ, Stewart SL, Kerlikowske K, Karliner LS. Disparities in abnormal mammogram follow-up time for Asian women compared with non-Hispanic white women and between Asian ethnic groups. Cancer 2017;123:3468–75. [PubMed: 28603859]
30. Press R, Carrasquillo O, Sciacca RR, Giardina EG. Racial/ethnic disparities in time to follow-up after an abnormal mammogram. J Womens Health (Larchmt) 2008;17:923–30. [PubMed: 18554094]
31. Sprague BL, Arao RF, Miglioretti DL, et al. National performance benchmarks for modern diagnostic digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology 2017;283:59–69. [PubMed: 28244803]
32. Dontchos BN, Narayan AK, Seidler M, et al. Impact of a same-day breast biopsy program on disparities in time to biopsy. J Am Coll Radiol 2019;16:1554–60. [PubMed: 31152690]
33. Ghate SV, Soo MS, Baker JA, Walsh R, Gimenez EI, Rosen EL. Comparison of recall and cancer detection rates for immediate versus batch interpretation of screening mammograms. Radiology 2005;235: 31–5. [PubMed: 15798165]
34. Stewart KA, Neumann PJ, Fletcher SW, Barton MB. The effect of immediate reading of screening mammograms on medical care utilization and costs after false-positive mammograms. Health Serv Res 2007;42:1464–82. [PubMed: 17610433]
35. Winkler NS, Freer P, Anzai Y, Hu N, Stein M. Impact of immediate interpretation of screening tomosynthesis mammography on performance metrics. Acad Radiol 2019;26:210–4. [PubMed: 29748047]
36. Raza S, Rosen MP, Chorny K, Mehta TS, Hulka CA, Baum JK. Patient expectations and costs of immediate reporting of screening mammography: talk isn't cheap. AJR Am J Roentgenol 2001;177:579–83. [PubMed: 11517050]
37. Schwartz LM, Woloshin S, Fowler FJ Jr, Welch HG. Enthusiasm for cancer screening in the United States. JAMA 2004;291:71–8. [PubMed: 14709578]
38. Roubidoux MA, Shih-Pei Wu P, Nolte ELR, Begay JA, Joe AI. Availability of prior mammograms affects incomplete report rates in mobile screening mammography. Breast Cancer Res Treat 2018;171: 667–73. [PubMed: 29951970]
39. Hickman SE, Woitek R, Le EPV, et al. Machine learning for workflow applications in screening mammography: systematic review and meta-analysis. Radiology 2022;302:88–104. [PubMed: 34665034]
40. Kyono T, Gilbert FJ, van der Schaar M. Improving workflow efficiency for mammography using machine learning. J Am Coll Radiol 2020;17:56–63. [PubMed: 31153798]
41. Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A deep learning model to triage screening mammograms: a simulation study. Radiology 2019;293:38–46. [PubMed: 31385754]
42. Miglioretti DL, Bissell MCS, Kerlikowske K, et al. Assessment of a risk-based approach for triaging mammography examinations during periods of reduced capacity. JAMA Netw Open 2021;4:e211974. [PubMed: 33764423]

TAKE-HOME POINTS

- Scheduling of the 4.9% of women with baseline mammograms for immediate mammographic interpretation and diagnostic evaluation could avert 11.8% of recalls for diagnostic imaging at a follow-up visit.
- Additional scheduling of the 4.3% of women with ≥5 years since prior mammogram, for a total of 9.2% of mammograms, could avert 19.2% of recalls for diagnostic imaging at a follow-up visit.
- Facilities with capacity to offer immediate interpretation to 25% of mammograms and preference for a simple rule could additionally schedule women <50 years of age with 1 to 4 years since prior mammogram to account for 40% of recalls, thereby averting a substantial portion of women being recalled for diagnostic imaging at a follow-up visit.
- Our prioritization approach may be applied regardless of the facility's profit status, location, screening volume, practice type, and academic affiliation.
- Immediate interpretation and same-day diagnostic workup could decrease loss to follow-up among women who are less likely to return for diagnostic workup after abnormal screening mammographic findings; however, if facilities serving marginalized communities are unable to offer this service, disparities could widen.

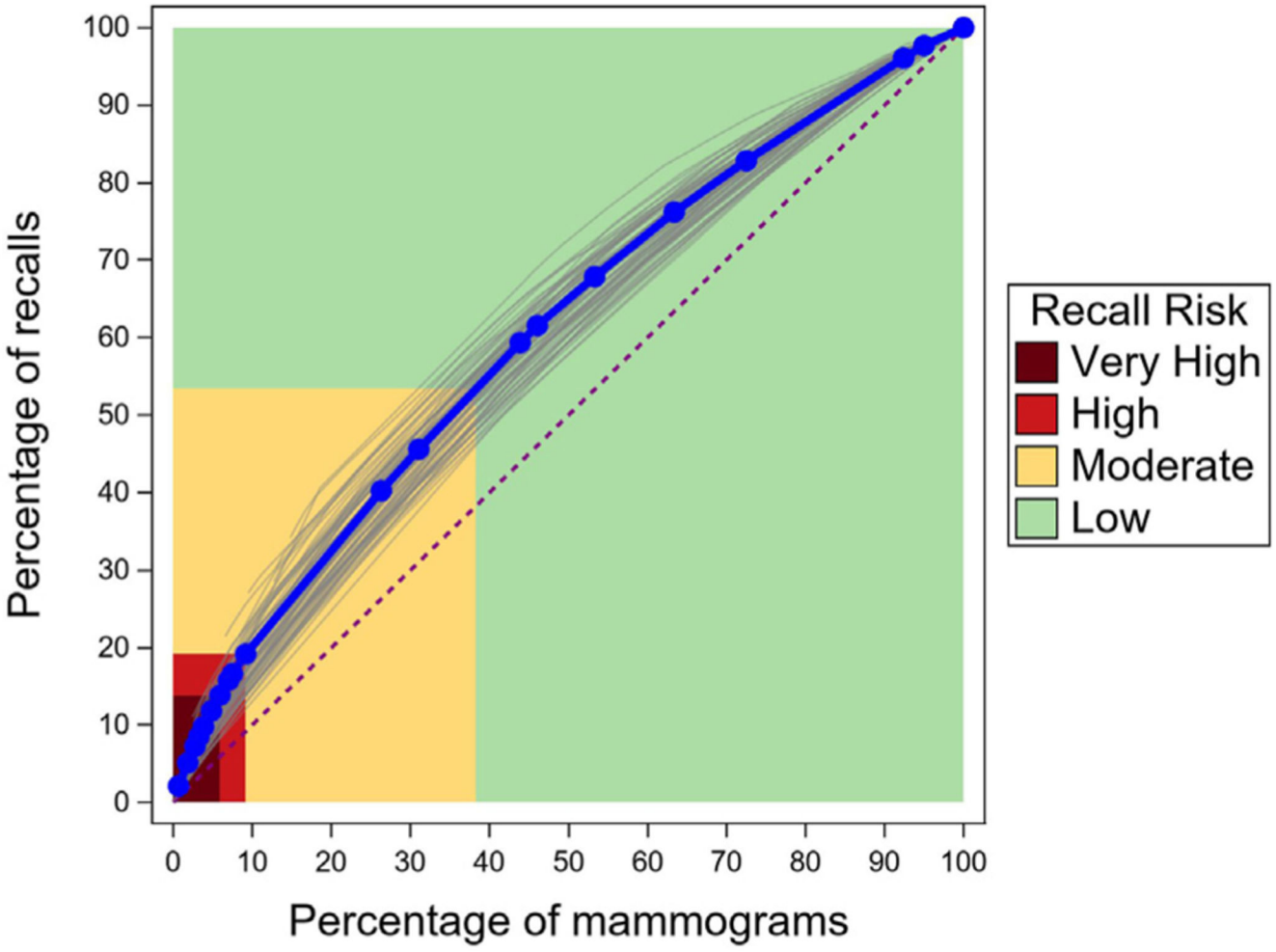

Fig. 1.

Cumulative percentile of recalls by cumulative percentile of mammograms from 19 subgroups identified by the classification tree models sorted from highest to lowest recall rate. The solid blue line represents the overall recall rate, with circles representing the efficiency ratios from the 19 subgroups identified by the classification tree models sorted from highest to lowest recall rate. The transparent gray lines represent facility-specific overall cumulative recall rates. The dashed 45° line represents the cumulative percentage of recalls expected using a random, non-risk-based approach.

Table 1.

Characteristics of 2,674,051 screening mammograms from 925,777 women in the study cohort by recall status, recall rate per 100 examinations, and efficiency ratio

Characteristic	Screening Mammograms		Recalls			
	n	%*	n	%*	Recall Rate per 100 (95% CI)	Efficiency Ratio†
Total	2,674,051	—	235,569	—	8.8 (8.3–9.4)	1.0
Age						
<40 y	24,130	0.9	4,122	1.7	17.1 (15.5–18.8)	1.9
40–49 y	540,562	20.2	67,604	28.7	12.5 (11.8–13.3)	1.4
50–59 y	835,280	31.2	72,642	30.8	8.7 (8.2–9.3)	1.0
60–69 y	795,024	29.7	58,757	24.9	7.4 (6.9–7.9)	0.8
≥70 y	479,055	17.9	32,444	13.8	6.8 (6.3–7.3)	0.8
Race/ethnicity						
Asian	273,942	10.5	21,388	9.3	7.8 (6.9–8.9)	0.9
Black	312,156	12.0	30,113	13.1	9.6 (8.5–10.9)	1.1
Hispanic	161,933	6.2	14,888	6.5	9.2 (8.6–9.8)	1.0
White	1,797,052	69.1	157,252	68.7	8.8 (8.2–9.4)	1.0
Other race or multiracial‡	56,331	2.2	5,395	2.4	9.6 (9.0–10.2)	1.1
Missing (%)*	72,637	2.7	6,533	2.8	9.0 (7.7–10.5)	1.0
First-degree family history of breast cancer						
No	2,178,701	81.5	191,875	81.5	8.8 (8.3–9.4)	1.0
Yes	495,350	18.5	43,694	18.5	8.8 (8.2–9.4)	1.0
History of false-positive recall						
No	2,227,105	83.3	189,425	80.4	8.5 (8.0–9.1)	1.0
Yes	446,946	16.7	46,144	19.6	10.3 (9.7–11.0)	1.2
History of benign biopsy result						
No	2,105,511	78.7	182,113	77.3	8.6 (8.1–9.2)	1.0
Yes	568,540	21.3	53,456	22.7	9.4 (8.8–10.1)	1.1
History of high-risk lesion§						
No	2,660,990	99.5	234,149	99.4	8.8 (8.3–9.4)	1.0

Characteristic	Screening Mammograms		Recalls			
	n	%*	n	%*	Recall Rate per 100 (95% CI)	Efficiency Ratio†
Yes	13,061	0.5	1,420	0.6	10.9 (9.8–12.1)	1.2
BI-RADS breast density						
Almost entirely fat	274,936	10.3	15,901	6.8	5.8 (5.3–6.3)	0.7
Scattered fibroglandular densities	1,165,732	43.6	94,434	40.1	8.1 (7.6–8.7)	0.9
Heterogeneously dense	1,025,523	38.4	104,806	44.5	10.2 (9.6–10.9)	1.2
Extremely dense	207,860	7.8	20,428	8.7	9.8 (9.1–10.6)	1.1
Time since last mammogram						
First mammogram	130,064	4.9	27,716	11.8	21.3 (19.6–23.2)	2.4
1 y	1,685,476	63.0	122,657	52.1	7.3 (6.8–7.8)	0.8
2 y	525,061	19.6	44,389	18.8	8.5 (7.9–9.0)	1.0
3 y	154,144	5.8	15,774	6.7	10.2 (9.6–10.8)	1.2
4 y	64,148	2.4	7,590	3.2	11.8 (11.2–12.5)	1.3
≥5 y	115,158	4.3	17,443	7.4	15.1 (14.3–16.1)	1.7
Screening modality						
Digital (2-D only)	1,816,553	67.9	165,385	70.2	9.1 (8.5–9.7)	1.0
Tomosynthesis	857,498	32.1	70,184	29.8	8.2 (7.5–9.0)	0.9

Note: CI = confidence interval.

*Column percentage among nonmissing, except for the missing category, which is percentage of total.

†Efficiency ratio is the percentage of recalls observed divided by the percentage of mammograms performed.

‡Other race or multiracial category included Native American, Alaskan Native, non-Hispanic with two or more reported races, and other.

§Atypical hyperplasia or lobular carcinoma in situ.

Table 2.

Characteristics of the 107 facilities and 2,673,890 screening mammograms by recall status in the study cohort and recall rate per 100 examinations

Characteristic	Facilities		Screening Mammograms		Recalls		Recall Rate per 100 (95% CI)
	n	%	n	%	n	%	
Total	107	—	2,673,890	—	235,473	—	8.8 (8.3–9.4)
Profit status							
For profit	25	23.4	537,904	20.1	45,075	21.1	8.4 (7.5–9.4)
Not for profit	70	65.4	1,871,930	70.0	168,968	78.9	9.0 (8.4–9.7)
Not reported	12	11.2	264,056	9.9	21,430	9.1	8.1 (7.2–9.2)
Screening volume*							
<1,000	21	19.6	41,588	1.6	4,888	2.1	11.8 (9.5–14.5)
1,000–2,999	25	23.4	268,160	10.0	23,823	10.1	8.9 (7.6–10.4)
3,000–4,999	32	29.9	720,161	26.9	70,092	29.8	9.7 (8.7–10.9)
≥5,000	29	27.1	1,643,981	61.5	136,670	58.0	8.3 (7.8–8.9)
Location							
Located within hospital	54	50.5	1,662,610	62.2	146,235	62.1	8.8 (8.1–9.5)
Office, not located within hospital	47	43.9	905,262	33.9	80,044	34.0	8.8 (8.1–9.7)
Mobile van	3	2.8	907	0.0	96	0.0	10.6 (9.0–12.4)
Other	3	2.8	105,111	3.9	9,098	3.9	8.7 (7.8–9.6)
Radiology practice type							
Multispecialty breast center	22	20.6	935,290	35.0	81,919	34.8	8.8 (8.1–9.5)
Full diagnostic radiology practice	69	64.5	1,414,660	52.9	124,991	53.1	8.8 (8.1–9.7)
Radiology practice limited to breast imaging	15	14.0	304,223	11.4	27,116	11.5	8.9 (7.6–10.4)
Screening only facility	1	0.9	19,717	0.7	1,447	0.6	7.3 (7.3–7.3)
Academic affiliation							
Academic	14	13.1	437,282	16.4	33,231	14.1	7.6 (6.7–8.6)
Nonacademic	93	86.9	2,236,608	83.6	202,242	85.9	9.0 (8.5–9.7)

Note: 161 mammograms from 19 facilities were excluded from the primary analysis because of missing facility-level covariates of interest. CI = confidence interval.

*Screening volume was computed as median annual screening volume at each facility and categorized into approximate quartiles.

Table 3.

Percentage of mammograms, percentage of recalls, recall rate, and cumulative efficiency ratio by risk group characterized by screening interval, age, history of false-positive recall, and breast density

Risk Group (% of Mammograms, % of Recalls)	Number (%) of Mammograms (n = 2,674,051)	Cumulative Percentage of Mammograms	Number (%) of Recalls (n = 235,569)	Cumulative Percentage of Recalls	Time Since Last Mammogram	Age (y)	Prior False-Positive Recall	Breast Density	Recall Rate per 100 (95% CI)	Cumulative Efficiency Ratio*
Very high risk (4.9%, 11.8%)	130,064 (4.9)	4.9	27,716 (11.8)	11.8	First mammogram	Any	Any	Any	21.3 (19.6–23.2)	2.4
High risk (4.3%, 7.4%)	115,158 (4.3)	9.2	17,443 (7.4)	19.2	≥5 y	Any	Any	Any	15.1 (14.3–16.1)	2.1
Moderate risk (29.1%, 34.2%)	296,517 (11.1)	20.3	34,803 (14.8)	33.9	1–4 y	<50	Any	Dense†	11.7 (11.0–12.5)	1.7
	139,143 (5.2)	25.5	13,511 (5.7)	39.7	1–4 y	<50	Any	Scattered	9.7 (8.9–10.5)	1.6
	342,169 (12.8)	38.3	32,343 (13.7)	53.4	1–4 y	≥50	Yes	Any	9.5 (8.9–10.1)	1.4
Low risk (61.7%, 46.6%)	525,180 (19.6)	57.9	41,532 (17.6)	71.0	2–4 y	≥50	Any	Any	7.9 (7.4–8.4)	1.2
	1,104,115 (41.3)	99.2	66,946 (28.4)	99.5	1 y	≥50	Any	Any	6.1 (5.7–6.5)	1.0
	21,705 (0.8)	100.0	1,275 (0.5)	100.0	1–4 y	<50	Any	Fatty	5.9 (5.2–6.6)	1.0

Note: Family history of breast cancer was not selected in the models, and screening modality groups are combined above and thus not shown in the table. CI = confidence interval.

*Cumulative efficiency ratio is the percentage of recalls observed divided by the percentage of mammograms performed when ordered from highest to lowest recall rate.

†Dense is heterogeneously dense or extremely dense.