

Breast Imaging

Calling all calcifications: a retrospective case control study

Anand K. Narayan^{*,1}, Delia M. Keating, Elizabeth A. Morris, Victoria L. Mango

Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

ARTICLE INFO

Keywords:

Screening
Mammography
Calcifications
Recalls
Breast cancer
Biopsy

ABSTRACT

Introduction: The majority of recalls from screening mammography do not result in biopsy recommendation. The purpose of this study was to evaluate if calcifications recalled from screening mammography are more likely to result in biopsy recommendations than other findings.

Methods: IRB-approved electronic medical record search was performed to obtain a random sample of screening mammograms assigned BI-RADS 0 assessment during 2014–2015. Primary reason for recall was classified as mass, asymmetry, focal asymmetry, calcifications, or distortion. Primary outcome was biopsy performed after diagnostic work-up. Secondary outcome was proportion of biopsies performed that were positive for cancer, positive predictive value 3 (PPV3). Logistic regression was used to compare reasons for recall (calcifications vs other findings) with biopsy recommendation proportions.

Results: Random database sampling yielded 402 screening examinations with BI-RADS 0 assessments with 449 total findings. Reasons for recall included calcifications (14.0%, 63/449), masses (15.8%, 71/449), asymmetries (50.8%, 228/449), focal asymmetries (14.3%, 64/449) and architectural distortions (5.1%, 23/449). Overall, 21.6% of recalls led to image-guided biopsy (87/402). Recalls for calcifications were more likely to result in biopsy compared with other types of findings (Adjusted OR 8.56, 95% CI 4.58 to 16.0, $p < 0.001$). No statistically significant differences were found in PPV3 proportions between calcification and non-calcification findings ($p = 0.812$).

Conclusion: Recalls for calcifications are much more likely to undergo biopsy compared with other findings. Increased biopsy rates for calcifications should be considered when recalling a patient from mammography screening in the context of practice specific positive predictive values and cancer detection rates.

1. Introduction

The majority of recalls from screening mammography do not result in biopsy recommendation [1]. Patients receive diagnostic workups tailored for specific recalled findings (e.g. calcifications, focal asymmetries, asymmetries, architectural distortions, masses) however these workups have variable potential to rule out malignancy. Prior studies have evaluated the positive predictive values of different mammographic findings [2], however there is limited information about which of these findings are more or less likely to result in biopsy recommendations. We hypothesize that calcifications recalled from screening mammography will be more likely to undergo biopsy compared with other findings.

2. Materials and methods

Study reporting followed STROBE guidelines for reporting observational studies [3].

2.1. Study design

An IRB approved, HIPAA compliant retrospective case control study was conducted to evaluate the hypothesis that patients recalled for calcifications were more likely to undergo biopsy compared with other findings (asymmetries, focal asymmetries, masses, architectural distortions). Cases included screening mammography patients recalled for calcification findings and controls included patients recalled for other findings.

* Corresponding author.

E-mail addresses: aknarayan@mg.harvard.edu (A.K. Narayan), keatingd@mskcc.org (D.M. Keating), morrise@mskcc.org (E.A. Morris), mangov@mskcc.org (V.L. Mango).

¹ Present address: Massachusetts General Hospital, 55 Fruit St. Wang 219H, Boston, MA, USA 02114.

2.2. Setting

Study was conducted at a single tertiary care urban cancer center.

2.3. Participants

Study included a random sample of women older than 40 undergoing digital screening mammography between 2014 and 2015 who were given BI-RADS 0 assessment on screening mammogram. To minimize biases associated with sample selection, none of the study authors was involved in selecting the random sample. We asked our data managers to derive a random sample of cases for our study using a random number generator. Patients who initially presented for screening examinations who described focal symptoms on the date of examination were given 0 BI-RADS ratings on the screening exam and told to return at a later date for diagnostic workup were excluded from our study.

2.4. Variables

Primary outcome was the proportion of women who underwent image guided biopsy after being recalled from screening mammography (BI-RADS 0). Secondary outcome was the proportion of biopsies actually performed which were positive for cancer (ductal carcinoma in situ or invasive carcinoma) (PPV3).

Using stated findings and the impression on screening mammography report, primary reason for recall was classified as mass, asymmetry, focal asymmetry, calcifications, or architectural distortion. The primary exposure evaluated was whether or not a screening examination was recalled for calcifications compared with other reasons for recall from screening examination (yes/no).

A priori we hypothesized the following variables to be potential confounders: age, race/ethnicity, history of prior cancer, history of prior breast biopsy (yes/no), number of prior breast biopsies, number of first degree relatives (mother, sister, daughters) with breast cancer, and breast density (almost entirely fatty, scattered fibroglandular densities, heterogeneously dense, extremely dense).

Hypothesized effect modifiers include whether or not screening mammogram was performed using tomosynthesis.

2.5. Data sources/measurement

Baseline demographics, screening and diagnostic mammography reports and outcomes were obtained from the electronic medical record. Baseline demographic information and risk factors were obtained from a data collection form obtained from every woman undergoing screening mammography with breast density categories obtained from the index screening mammogram report. Follow up days were calculated from the date of the index screening mammogram to the date of the patient's most recent screening or diagnostic mammogram.

2.6. Study size

We hypothesized that calcification findings would lead to a higher proportion of biopsy recommendations compared with non-calcification findings. Assuming that 50% of calcification recalls would lead to biopsy compared with 25% of non-calcification recalls leading to biopsy, this would require a total sample size of 110 individuals for a power level of 80%. To evaluate our secondary hypothesis, prior population based estimates of PPV3 estimate benchmark PPV3 proportions as 31% [4]. If we assume that calcifications' PPV3 proportion would be approximately half as other findings (15%), we would need a total of 350 patients (assuming that approximately 1/5 of cases are recalled for calcifications) for a statistical power of 80%.

2.7. Quantitative variables

Logistic regression analyses were performed comparing primary and secondary outcomes with the primary exposure (recall for calcifications), unadjusted and adjusted for potential confounders. Effect modification by screening mammography technique (2D vs 3D) was evaluated using interaction terms. p values < 0.05 were considered statistically significant. Analyses were conducted using STATA 11 (StataCorp, College Station, TX).

3. Results

A total of 450 patients randomly selected from 2014 to 2015 with BI-RADS 0 assessments. Further review of indications for presentation revealed that 48 of these patients presented with focal symptoms leading to diagnostic evaluations. Exclusion of these patients yielded 402 patients undergoing screening examinations with BI-RADS 0 assessments with 449 total findings (31 of the 402 patients had multiple findings that were called back from screening mammography). 8.5% of patients underwent tomosynthesis (34/402). Reasons for recall included calcifications (14.0%, 63/449), masses (15.8%, 71/449), asymmetries (50.8%, 228/449), focal asymmetries (14.3%, 64/449) and architectural distortions (5.1%, 23/449).

At baseline, patients recalled for calcifications ($n = 63$ patients) demonstrated mostly similar characteristics compared with patients not recalled for calcifications ($n = 339$ patients) in term of demographics and breast cancer risk factors (Table 1). Patients recalled for calcification findings were borderline statistically significantly more likely to present with history of prior breast biopsy (39.7 vs 27.3%, $p = 0.047$). All patients presented for diagnostic mammography at our institution after initial screening mammogram with BI-RADS 0 assessment. No statistically significant differences in mean follow up (367 vs 369 days, $p = 0.978$) were found comparing patients recalled for calcification versus non-calcification findings.

Overall, 21.6% of recalled patients underwent image-guided biopsy

Table 1

Baseline demographic characteristics comparing patients recalled for calcification findings versus patients recalled for non-calcification findings.

Parameter	Calcification recalls (n = 63)	Non-calcification recalls (n = 339)	p value
Age (mean)	57.1	56.4	0.634
Race/ethnicity (%)			
White	49.2	44.1	0.641
Black	25.4	29.3	
Hispanic	7.9	13.3	
Asian	7.9	5.6	
Other	9.5	7.7	
Breast density (%)			
Almost entirely fatty	9.5	9.4	0.181
Scattered fibroglandular densities	25.4	36.0	
Heterogeneously dense	50.8	47.2	
Extremely dense	14.3	7.4	
Number of first degree relatives with breast cancer (%)			
0	71.4	79.8	0.138
1	22.2	18.4	
2	4.8	1.5	
3	1.6	0.3	
Prior history of breast cancer (%)			
Yes	0.0	2.7	0.190
History of prior breast biopsy (%)			
Yes	39.7	27.3	0.047
Number of prior breast biopsies (mean)	0.54	0.39	0.187

Table 2
Unadjusted and adjusted odds ratios of undergoing biopsy by specific finding for which patient was recalled from screening mammography.

Finding	Unadjusted	p value	Adjusted	p value
Calcifications	7.53 (4.21, 13.46)	< 0.001	8.56 (4.58, 16.02)	< 0.001
Focal asymmetry	0.91 (0.47, 1.76)	0.778	0.97 (0.48, 1.94)	0.921
Architectural distortion	0.75 (0.25, 2.27)	0.611	0.58 (0.19, 1.82)	0.350
Mass	1.55 (0.86, 2.77)	0.143	1.67 (0.91, 3.08)	0.097
Asymmetry	0.33 (0.20, 0.54)	< 0.001	0.30 (0.18, 0.50)	< 0.001

Biopsy Percentages by Finding

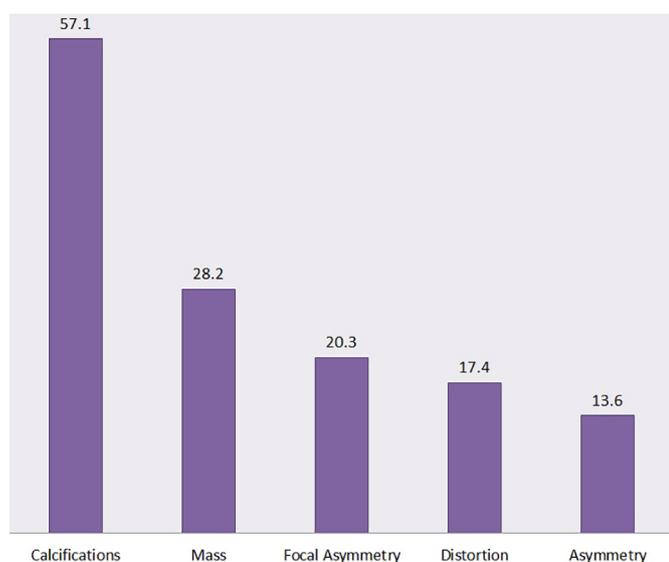


Fig. 1. Biopsy percentages by finding.

Figure describes the percentage of patients who underwent biopsy for each type of finding.

(87/402) leading to the diagnosis of 23 malignancies (17 invasive, 6 in situ), corresponding to an overall PPV3 of 26.4%. Recalls for calcifications were more likely to result in biopsy compared with other types of findings (Adjusted OR 8.56, 95% CI 4.58 to 16.0, $p < 0.001$) (Table 2). Biopsy for calcifications was most common (57.1%, 36/63) followed by mass (28.2%, 20/71), focal asymmetry (20.3%, 13/64), distortion (17.4%, 4/23), and asymmetry (13.6%, 31/228) (Fig. 1). PPV3 proportions were highest for masses (30.0%, 6/20) followed by calcifications (27.8%, 10/36), architectural distortions (25.0%, 1/4), focal asymmetries (23.1%, 3/13), and asymmetries (16.1%, 5/31). Higher proportions of in situ carcinomas were diagnosed as a result of calcification findings (50.0%, 5/10) compared with non-calcification findings (7.7%, 1/13). No statistically significant differences were found in PPV3 proportions between calcification and non-calcification findings (OR 1.12, 95% CI 0.38 to 3.26, $p = 0.812$). Effect modification analysis of tomosynthesis revealed no evidence of effect modification on the probability that calcification recalls would undergo biopsy ($p = 0.644$).

4. Discussion

Most findings recalled from screening mammography do not lead to biopsy; however recalls for calcifications are much more likely to undergo biopsy compared to other findings. These results suggest that caution should be exercised in recalling patients from screening

mammography for calcifications in light of the higher probability that calcification findings will lead to biopsy. Our results are similar to prior studies evaluating recalls from screening mammography [2,5].

Venkatesan et al. studied 10,262 women in the San Francisco Mammography Registry and found that asymmetries were associated with the lowest positive predictive values compared with masses, calcifications and architectural distortions, similar to our results. Although they did not specifically evaluate the proportion of each type of finding leading to biopsy recommendation, they found no statistically significant differences comparing PPVs from calcification findings compared with non-calcification findings. Similarly, Lipasti et al. found similar PPVs comparing calcification findings without associated masses (44.0%) compared with “tumor-like” masses, asymmetries and distortions (40.3%) [5]. The similarities in PPV proportions in prior studies as well as our study suggest that radiologists may be implicitly or explicitly weighing the possibility that calcification findings would be more likely to lead to biopsy compared with other findings at the time of recall. Similar PPV3 values for calcification versus non-calcification findings, which fall within established benchmarks, suggest a higher biopsy recommendation percentage is appropriate for recalled calcifications in our clinical practice. However, with considerable variation in PPV proportions nationwide [6], radiology practices may consider conducting finding specific audits to evaluate practice specific PPV3 of calcification recalls in light of evidence suggesting that calcifications are more likely to undergo biopsy after recall from screening mammography.

Our findings that calcification findings are more likely to result in biopsy are consistent with outcomes from diagnostic mammography. While masses, asymmetries and focal asymmetries often represent various benign findings (i.e. cysts, lymph nodes) or simply overlapping tissue, calcifications have a relatively smaller range of features (i.e. layering or vascular) on magnification mammography that lead to definitive characterization as benign [7]. In our study, architectural distortion was associated with a lower probability of biopsy compared with calcification findings. Although mammographic architectural distortion in general has been associated with high PPV values [8], architectural distortion detected on screening mammography may not persist on diagnostic (2D or 3D) views or may reflect histories of prior surgery (i.e. faded, unapparent skin scars).

Previous studies provide some suggestions for individual radiologists or radiology practices with high biopsy rates and low PPVs from calcification findings. Most of the previously described potential strategies involve either reducing recall rates or following calcifications (instead of biopsying them). In terms of reducing recall rates, several studies have highlighted the importance of obtaining prior mammograms when reviewing screening examinations. The availability of prior exams has been associated large reductions in recall rates ranging between 19 and 34% [9]. In addition to aggressively seeking prior examinations, radiologists can use the morphology and distribution of calcification findings to increase or decrease their level of suspicion. Using 5th edition BI-RADS terms, Kim et al. studied PPVs associated with different morphologies and distributions [10]. They found an overall PPV value of 22% for calcification findings however PPVs varied widely based on the morphology and distribution of the calcifications. In terms of morphology, amorphous calcifications demonstrated the lowest PPV values (7.9%), followed by coarse heterogeneous (17.8%), fine pleomorphic (63.2%), and fine linear/fine linear branching (100%) morphologies. In terms of distribution, regional calcifications demonstrated the lowest PPV values (8.8%), followed by grouped (14.3%), segmental (63.6%) and linear (87.5%) distributions. Evaluating combinations of calcification morphology and distribution, amorphous calcifications demonstrated PPV values of 4.2% in regional distributions and 7.6% with grouped distributions. Alternatively, coarse heterogeneous calcifications with linear or segmental distributions demonstrated PPV values of 50.0% while fine pleomorphic calcifications with linear or segmental distributions demonstrated PPV values of

93.8%. To further evaluate microcalcifications, Burnside et al. used mammographic features and breast cancer risk factors in Bayesian computer models and found that using these features would lead to a 34% reduction in biopsies and increase in PPV from 22 to 31% [11]. Importantly, these performance improvements could be achieved without missing any additional breast cancers. These studies suggest the possibility that radiology practices with high biopsy rates for calcification findings can use morphology, distribution and breast cancer risk factors to reduce unnecessary biopsies without sacrificing cancer detection.

Additionally, the possibility of following indeterminate calcifications seen on screening mammogram has been evaluated by newer technologies including breast MRI and contrast-enhanced mammography. Bennani-Baiti and Baltzer performed a meta-analysis of twenty studies evaluating the performance of breast MRI for the evaluation of micro calcifications [12]. They found that breast MRI had limited performance for calcifications given BI-RADS 3 and 5 categories however for BI-RADS 4 microcalcifications, breast MRI performed well with a negative likelihood ratio of 0.099 (95% CI: 0.07, 0.15) and a positive likelihood ratio of 5.03 (95% CI: 3.43, 7.53). They concluded that patients with BI-RADS 4 microcalcifications with pre-test probabilities up to 18.4% would benefit from breast MRI examinations to evaluate for the presence or absence of enhancement to rule out malignancy. Similarly Cheung et al. found that enhancement in contrast-enhanced mammography improved positive predictive values for amorphous and pleomorphic calcifications with negative predictive values of 95% [13]. With the acquisition of low energy images during contrast-enhanced mammography examinations, contrast enhanced mammography also offers the possibility of analyzing mammographic features of microcalcifications as well as contrast enhancement.

Limitations of our study include inherent limitations associated with retrospective studies, reader variability and changes in reading patterns associated with wider dissemination of tomosynthesis. Our study was limited by retrospective data collection. Although all of the patients in our study returned to our institution for diagnostic mammography after screening recall, patients may have received either biopsies or cancer diagnoses at other institutions. Though some fraction of patients may have presented to other institutions for biopsy after initial screening mammogram and diagnostic work up, mean follow up did not differ between patients recalled for calcifications versus patients recalled for non-calcification finding, limiting possibilities that loss to follow up would have influenced our conclusions comparing patients recalled for calcifications versus non-calcification findings. Additionally, our study did not evaluate individual reader performance and variability in recall rates may have influenced overall recall proportions. However > 10 radiologists interpreted screening mammograms during the study period making it unlikely that overall recall proportions are skewed by outliers and overall PPV3 values are within accepted practice benchmarks [14]. Additionally Venkatesan et al. noted similar mean and median PPVs for calcification findings, suggesting consistency in recall proportions for calcification findings. Finally, the increasing adoption of tomosynthesis may influence the applicability of our study results to future practice. In our study, most of the screening examinations were performed using 2D digital mammography. Recent studies on tomosynthesis suggest that tomosynthesis is associated with decreased numbers of false positive examinations, mostly due to decreased asymmetries and focal asymmetries requiring additional work up [15]. As the primary outcome of our study relates to calcifications, wider adoption of tomosynthesis may be found to have a greater influence on performance metrics of asymmetries and focal asymmetries with comparatively less impact on calcification findings. Performance of our study at a large cancer center with subspecialty breast radiologists may limit broad applicability of these results, further emphasizing the need for practice specific assessment.

5. Conclusions

Most findings recalled from screening mammography do not lead to biopsy; however recalls for calcifications are much more likely to undergo biopsy. Increased biopsy rates for calcifications should be considered when recalling a patient from screening in the context of practice specific positive predictive values and cancer detection rates.

Declarations of interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data statement

Our data is unavailable to access or post as the research data is confidential.

References

- [1] Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Ann Intern Med* Oct 18 2011;155(8):481–92.
- [2] Venkatesan A, Chu P, Kerlikowske K, Sickles EA, Smith-Bindman R. Positive predictive value of specific mammographic findings according to reader and patient variables. *Radiology* Mar 2009;250(3):648–57.
- [3] Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* Oct 16 2007;4(10):e297.
- [4] Sprague BL, Arao RF, Miglioretti DL, Henderson LM, Buist DS, Omega T, et al. National performance benchmarks for modern diagnostic digital mammography: update from the breast cancer surveillance consortium. *Radiology* Apr 2017;283(1):59–69.
- [5] Lipasti S, Anttila A, Pamilo M. Mammographic findings of women recalled for diagnostic work-up in digital versus screen-film mammography in a population-based screening program. *Acta Radiol* Jun 2010;51(5):491–7.
- [6] Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DS, Kerlikowske K, et al. National performance benchmarks for modern screening digital mammography: update from the breast cancer surveillance consortium. *Radiology* Apr 2017;283(1):49–58.
- [7] Wilkinson L, Thomas V, Sharma N. Microcalcification on mammography: approaches to interpretation and biopsy. *Br J Radiol* Jan 2017;90(1069):20160594.
- [8] Bahl M, Baker JA, Kinsey EN, Ghate SV. Architectural distortion on mammography: correlation with pathologic outcomes and predictors of malignancy. *AJR Am J Roentgenol* Dec 2015;205(6):1339–45.
- [9] Hakim CM, Catullo VJ, Chough DM, Ganott MA, Kelly AE, Shinde DD, et al. Effect of the availability of prior full-field digital mammography and digital breast tomosynthesis images on the interpretation of mammograms. *Radiology* Jul 2015;276(1):65–72.
- [10] Kim SY, Kim HY, Kim EK, Kim MJ, Moon HJ, Yoon JH. Evaluation of malignancy risk stratification of microcalcifications detected on mammography: a study based on the 5th edition of BI-RADS. *Ann Surg Oncol* Sep 2015;22(9):2895–901.
- [11] Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology* Sep 2006;240(3):666–73.
- [12] Bennani-Baiti B, Baltzer PA. MR imaging for diagnosis of malignancy in mammographic microcalcifications: a systematic review and meta-analysis. *Radiology* Jun 2017;283(3):692–701.
- [13] Cheung YC, Juan YH, Lin YC, Lo YF, Tsai HP, Ueng SH, et al. Dual-energy contrast-enhanced spectral mammography: enhancement analysis on BI-RADS 4 non-mass microcalcifications in screened women. *PLoS One* Sep 9 2016;11(9):e0162740.
- [14] Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, Carney PA, Kerlikowske K, Buist DS, Weaver DL, Barlow WE, Ballard-Barbash R. Performance benchmarks for screening mammography. *Radiology* 2006 Oct;241(1):55–66.
- [15] Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* Jun 25 2014;311(24):2499–507.