Factors Associated With False-Positive Recalls in Mammography Screening

Xinhe Mao, MD, MSc[1]; Wei He, PhD[1,2,3]; Keith Humphreys, PhD[1]; Mikael Eriksson, PhD[1]; Natalie Holowko, PhD[1,4]; Fredrik Strand, MD, PhD[5,6]; Per Hall, MD, PhD[1,7]; and Kamila Czene, PhD[1]

ABSTRACT

Background: We aimed to identify factors associated with false-positive recalls in mammography screening compared with women who were not recalled and those who received true-positive recalls. **Methods:** We included 29,129 women, aged 40 to 74 years, who participated in the Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA) between 2011 and 2013 with follow-up until the end of 2017. Nonmammographic factors were collected from questionnaires, mammographic factors were generated from mammograms, and genotypes were determined using the OncoArray or an Illumina custom array. By the use of conditional and regular logistic regression models, we investigated the association between breast cancer risk factors and risk models and false-positive recalls. **Results:** Women with a history of benign breast disease, high breast density, masses, microcalcifications, high Tyrer-Cuzick 10-year risk scores, KARMA 2-year risk scores, and polygenic risk scores were more likely to have mammography recalls, including both false-positive and true-positive recalls. Further analyses restricted to women who were recalled found that women with a history of benign breast disease and dense breasts had a similar risk of having false-positive and true-positive recalls, whereas women with masses, microcalcifications, high Tyrer-Cuzick 10-year risk scores, KARMA 2-year risk scores, and polygenic risk scores were more likely to have true-positive recalls than false-positive recalls. **Conclusions:** We found that risk factors associated with false-positive recalls were also likely, or even more likely, to be associated with true-positive recalls in mammography screening.

J Natl Compr Canc Netw 2023;21(2):143–152.e4
doi: 10.6004/jnccn.2022.7081

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [2]Chronic Disease Research Institute, the Children's Hospital, and National Clinical Research Center for Child Health, School of Public Health, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China; [3]Department of Nutrition and Food Hygiene, School of Public Health, Zhejiang University, Hangzhou, Zhejiang, China; [4]Department of Medicine Solna, Clinical Epidemiology Division, Karolinska Institutet, Stockholm, Sweden; [5]Department of Radiology, Karolinska University Hospital, Stockholm, Sweden; [6]Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden; and [7]Department of Oncology, Södersjukhuset, Stockholm, Sweden.

Background

Mammography screening reduces breast cancer mortality by an estimated 26% to 41%.[1,2] However, false-positive recalls—recalling women with abnormal mammograms who, on further testing, are not found to have breast cancer—can cause psychological burden.[3–6] Such burden may further decrease women's screening reattendance by undermining women's confidence in the benefits of mammography.[4,7,8] In Sweden, approximately 2.5 per 100 women attending mammography screening experience a false-positive recall at a single screening round.[9] Because Swedish women are screened every second year from aged 40 to 74 years, there is a high lifetime risk of having a false-positive mammography recall. Similarly, in the United States, it is estimated that 30% to 50% of women who participate in mammography screening will have a false-positive recall over a 10-year period.[10,11]

Although false-positive recalls cannot be eliminated, they can be minimized. Better understanding the association between breast cancer risk factors and false-positive recalls may help reduce their occurrence. Previous studies have found that high breast density is associated with false-positive recalls.[11–13] However, no study thus far has investigated the association of false-positive recalls with other mammographic features (eg, microcalcifications and masses) and breast cancer risk prediction models (eg, Tyrer-Cuzick model). Furthermore, false-positive recalls should be reduced, but not at the cost of missing true tumors. Therefore, when investigating determinants of false-positive recalls, the risk of true-positive recalls should be considered, but unfortunately this has been neglected in previous studies.[11,12]

Using the Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA), a population-based screening cohort, we investigated the association of mammographic features, nonmammographic features, and breast cancer risk prediction models with false-positive recalls compared with women who were not recalled as well as those who received a true-positive recall.


See JNCCN.org for supplemental online content.

Methods

Data Sources

The KARMA study comprises women who attended mammography screening or clinical mammography at 4 hospitals in Sweden between January 2011 and March 2013.[14] Blood samples were collected at baseline from 98% of KARMA participants. The study further collected data from questionnaires and mammograms. Detailed information about recruitment, participant characteristics, questionnaires, mammograms collection, and follow-up can be found elsewhere.[14]

Using the unique Swedish personal identification number,[15] we linked KARMA to the Stockholm-Gotland Breast Cancer Register, the National Quality Register for Breast Cancer, and the Stockholm mammography screening program. This program invites all women in Stockholm aged 40 to 74 years for mammography screening at 18- to 24-month intervals.[9,16–18]

Study Population

We identified all 32,185 KARMA participants who participated in the Stockholm mammography screening program. We then excluded women who had breast enlargement (n=713), breast reduction (n=975), or other breast surgeries (n=816) and those who were diagnosed with breast cancer before the KARMA baseline (n=552), leaving 29,129 for the final analyses (supplemental eFigure 1, available with this article at JNCCN.org).

For analysis of recall rates, we included all mammography screening records within 30 days of KARMA recruitment (n=28,192). Among these screenings, there were 796 recalls. To examine associations of risk factors with both false-positive and true-positive recalls, we conducted a matched case-control study based on screenings performed between 2011 and 2015 for the 29,129 women (supplemental eFigure 1). Specifically, we identified 1,550 women who received their first mammography recall at or after entering the KARMA cohort. Selecting recalls in this way increased our study's statistical power (compared with using only recalls at enrollment), given that most women attended >1 screening during the study period. For each recalled woman, we randomly selected 5 age-matched (±1 year) and screening history–matched control individuals (women who were not recalled at the time and had never been recalled previously). We then further categorized recalled women as having a false-positive (n=1,233) or true-positive recall (n=317).

In Stockholm, all screening units have used double-reading with a consensus decision method, which involves 2 radiologists independently assessing mammograms for each participant to decide whether the woman is "healthy" or needs to be recalled for further assessment.[9] Each radiologist decides whether a woman should be recalled or not predominantly based on any suspicious mammographic findings, such as masses, microcalcifications, and asymmetry of density, while reviewing prior mammograms (for women at second or later screens) for comparison. If either radiologist notes a suspicious finding, the case is discussed until a consensus is reached. Information on family history of breast cancer or other breast cancer risk factors is not collected or considered at mammography clinics. In this study, false-positive recalls were defined as not being diagnosed with breast cancer (including invasive and in situ breast cancer) between the date of being recalled and the next scheduled screening visit. True-positive recalls were defined as being diagnosed with breast cancer within the same period. To categorize the screening outcomes, we used data from the Swedish Cancer Register with follow-up until the end of 2017.

Nonmammographic Features

Information on the following breast cancer risk factors was retrieved from the KARMA questionnaire[14]: years of education (<10, 10–12, >12 years), family history of breast cancer (no, yes), history of benign breast disease (no, yes), age at menarche (<14, ≥14 years), nulliparity (no, yes), number of children (0, 1 or 2, >2), age at first birth (<25, 25–35, >35 years), duration of breastfeeding (0, <6, 6–12, >12 months), use of oral contraceptive (never, ever), use of any hormone replacement therapy (never, ever), body mass index (<25, 25.0–29.9, ≥30.0 kg/m^2), physical activity (<40, 40–44.9, ≥45 metabolic equivalents of task h/d), smoking status (never, ever), and alcohol consumption (0, 0.1–10, >10 g/d).

Mammographic Features

Mammograms of both breasts were used to measure mammographic features. Dense area was measured for each breast using the automated STRATUS method.[19] The number of masses and microcalcification clusters was measured using FDA-approved computer-aided detection software (M-Vu CAD; iCAD).[20] Mammographic features were categorized as dense area (<9, 9 to <20, >20 cm^2), number of masses (0, ≥1), and microcalcifications (0, ≥1). For a woman with a recall (and her subsequent matching control individuals), we defined the recalled (right or left) side of the breast as the "recalled side" and the other as the "contralateral side." Asymmetry of mammographic features between the 2 breasts was defined using the difference between dense area and the number of masses and microcalcification clusters on the recalled side compared with the contralateral side. Equal was defined as within 6 cm^2 of dense area, same number of masses, and same number of microcalcification clusters, respectively.

Table 1. Association Between Breast Cancer Risk Factors, Risk Models, and Mammography Screening Recall Rates

		Recalls (Crude rate per 1,000 screenings)		
	Number of Screenings	Yes	No	*P* Value[a]
Overall	28,192	796 (28.3)	27,396 (971.7)	
Nonmammographic factors				
Age at mammogram				.004
40–49 y	10,684	345 (32.3)	10,339 (967.7)	
50–59 y	8,163	201 (24.6)	7,962 (975.4)	
60–74 y	9,345	250 (26.8)	9,095 (973.2)	
History of benign breast disease				<.001
No	21,612	540 (25.0)	21,072 (975.0)	
Yes	6,038	235 (38.9)	5,803 (961.1)	
Family history of breast cancer				.002
No	23,565	641 (27.2)	22,924 (972.8)	
Yes	3,805	137 (36.0)	3,668 (964.0)	
Years of education				.140
≤9	2,430	64 (26.3)	2,366 (973.7)	
10–12	9,128	234 (25.6)	8,894 (974.4)	
>12	15,646	466 (29.8)	15,180 (970.2)	
Age at menarche				.136
<14 y	17,727	522 (29.4)	17,205 (970.6)	
≥14 y	9,834	259 (26.3)	9,575 (973.7)	
Nulliparity				.494
No	23,811	665 (27.9)	23,146 (972.1)	
Yes	4,329	129 (29.8)	4,200 (970.2)	
Number of children				.371
0	4,329	129 (29.8)	4,200 (970.2)	
1 or 2	17,477	502 (28.7)	16,975 (971.3)	
>2	6,334	163 (25.7)	6,171 (974.3)	
Age at first birth[b]				.537
<25 y	7,377	207 (28.1)	7,170 (971.9)	
25–35 y	14,276	406 (28.4)	13,870 (971.6)	
>35 y	2,149	52 (24.2)	2,097 (975.8)	
Breastfeeding duration,[b] mo				.468
0	504	12 (23.8)	492 (976.2)	
<6	3,574	98 (27.4)	3,476 (972.6)	
6–12	7,545	197 (26.1)	7,348 (973.9)	
>12	12,045	357 (29.6)	11,688 (970.4)	
BMI, kg/m^2				.335
<25	16,183	442 (27.3)	15,741 (972.7)	
25–29.9	8,523	242 (28.4)	8,281 (971.6)	
≥30	3,350	107 (31.9)	3,243 (968.1)	
Oral contraceptive				.283
Never used	5,595	170 (30.4)	5,425 (969.6)	
Ever used	22,578	626 (27.7)	21,952 (972.3)	

(*continued on next page*)

Table 1. Association Between Breast Cancer Risk Factors, Risk Models, and Mammography Screening Recall Rates (cont.)

	Number of Screenings	Recalls (Crude rate per 1,000 screenings)		*P* Value[a]
		Yes	No	
Nonmammographic factors (cont.)				
Hormone replacement therapy[c]				.810
Never used	9,176	220 (24.0)	8,956 (976.0)	
Ever used	6,634	163 (24.6)	6,471 (975.4)	
Physical activity, MET h/d				.299
<40	11,030	293 (26.6)	10,737 (973.4)	
40–44.9	9,723	279 (28.7)	9,444 (971.3)	
≥45	6,899	210 (30.4)	6,689 (969.6)	
Smoking status				.556
Never	12,642	349 (27.6)	12,293 (972.4)	
Ever	15,465	445 (28.8)	15,020 (971.2)	
Alcohol consumption, g/d				.477
0	4,858	140 (28.8)	4,718 (971.2)	
0.1–10	17,122	497 (29.0)	16,625 (971.0)	
>10	5,990	156 (26.0)	5,834 (974.0)	
Mammographic factors				
Dense area,[d] cm^2				<.001
<9	5,279	109 (20.6)	5,170 (979.4)	
9 to <20	4,785	123 (25.7)	4,662 (974.3)	
≥20	15,157	501 (33.1)	14,656 (966.9)	
Masses[d]				<.001
0	12,173	209 (17.2)	11,964 (982.8)	
≥1	9,241	394 (42.6)	8,847 (957.4)	
Microcalcifications[d]				<.001
0	18,912	482 (25.5)	18,430 (974.5)	
≥1	2,502	121 (48.4)	2,381 (951.6)	
Breast cancer risk prediction models				
Tyrer-Cuzick 10-y risk scores				.079
<3%	14,399	399 (27.7)	14,000 (972.3)	
3% to <5%	9,841	264 (26.8)	9,577 (973.2)	
≥5%	3,952	133 (33.7)	3,819 (966.3)	
KARMA 2-y risk scores				<.001
<0.6%	16,718	360 (21.5)	16,358 (978.5)	
0.6% to <1.0%	2,285	78 (34.1)	2,207 (965.9)	
≥1.0%	2,341	120 (51.3)	2,221 (948.7)	
PRS quartiles[e]				<.001
Q1	2,018	51 (25.3)	1,967 (974.7)	
Q2	2,048	73 (35.6)	1,975 (964.4)	
Q3	2,063	75 (36.4)	1,988 (963.6)	
Q4	2,130	148 (69.5)	1,982 (930.5)	

Column totals may not equal the total number of subjects because of missing values.
Abbreviations: BMI, body mass index; KARMA, Karolinska Mammography Project for Risk Prediction of Breast Cancer; MET, metabolic equivalents of task; PRS, polygenic risk scores; Q, quartile.
[a]*P* values derived from chi-square tests.
[b]Among parous women.
[c]Among postmenopausal women.
[d]Mammographic features of the recalled breast.
[e]PRS quartile cutoffs were defined based on data from women who were not recalled.

Breast Cancer Risk Models

The Tyrer-Cuzick 10-year breast cancer risk score was categorized as low (<3%), medium (3% to <5%), or high (≥5%).[14,21] The KARMA 2-year risk score was categorized as low (<0.6%), medium (0.6% to <1.0%), or high (≥1.0%).[20] A weighted breast cancer polygenic risk score for each genotyped individual with European ancestry was calculated by including 313 single-nucleotide polymorphisms that have previously been identified.[22]

Statistical Analyses

We estimated the mammography screening recall rates (numbers per 1,000 screening) by nonmammographic features, mammographic features, and breast cancer risk models. Chi-square tests ($P<.1$) were used to examine and select risk factors that differed by recall status. We then used conditional logistic regression models to investigate the association between these selected breast cancer risk factors and both false-positive and true-positive recalls, respectively, compared with age-matched control individuals. Furthermore, we also conducted logistic regression to directly compare women with false-positive recalls with those with true-positive recalls. Additional analyses were also conducted for factors ($P\geq.1$ in Table 1) that were not associated with recalls.

All analyses were performed using SAS 9.4 (SAS Institute Inc.) and R version 3.6 (R Foundation for Statistical Computing). All *P* values were 2-sided. The Regional Ethical Review Board in Stockholm, Sweden, approved the study.

Results

For every 1,000 mammography screenings, there were, on average, 28 women recalled for further examinations (Table 1). Women with the following breast cancer risk factors had significantly higher recall rates than those without: family history of breast cancer, history of benign breast disease, dense breasts (dense area ≥20 cm^2), masses, microcalcifications, and asymmetry of these 3 mammographic features (more prevalent in the recalled breast) (Table 1 and supplemental eTable 1). In addition, women with high breast cancer risk, measured by all 3 risk models (Tyrer-Cuzick, KARMA, and polygenic models), also had higher recall rates than those with low risk (Table 1).

In contrast, age was the only breast cancer risk factor negatively associated with recall rates (Table 1). Specifically, we found that false-positive recall rates were highest in women aged 40 to 49 years and declined with age, whereas true-positive recall rates increased with age (Figure 1). Of note, the number of false-positive recalls per true-positive recall was 9.78 (95% CI, 7.11–15.09) in women aged 40 to 49 years, which was 4 times higher than the 1.78 (95% CI, 1.38–2.33) observed among women aged 60 to 74 years (Figure 1).

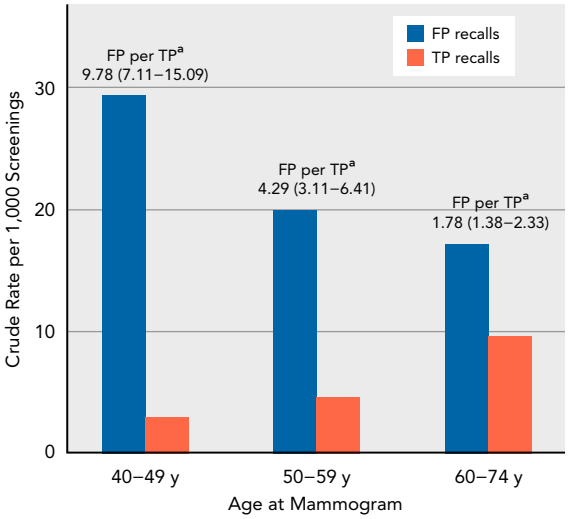


Figure 1. False-positive (FP) and true-positive (TP) recall rates, by age at screening.
[a]Number of FP recalls per one TP recall presented with 95% confidence intervals.

Nonmammographic Factors

Women with a history of benign breast disease were more likely than those without to have mammography recalls (Table 1), including false-positive and true-positive recalls (Table 2). Further analyses restricted to women who were recalled found that those with and without a history of benign breast disease had a similar risk of having a false-positive and true-positive recall (Figure 2). Women with a family history of breast cancer were more likely than those without to have a recalled mammogram (Table 1), particularly for true-positive recalls (Table 2). Further analyses restricted to women who were recalled found that those with a family history of breast cancer were more likely to have a true-positive recall than false-positive recall (Figure 2). None of the other factors favor false-positive over true-positive recalls (supplemental eFigure 2).

Mammographic Factors

Women with high breast density, masses, and microcalcifications were more likely than those without to have a recalled mammogram (Table 1), including both false-positive and true-positive recalls (Table 2). Further analyses restricted to women who were recalled found that those with masses and microcalcifications were more likely to have a true-positive than a false-positive recall, whereas no significant difference was found for breast density (Figure 2). Furthermore, women with asymmetric breast features were also more likely to have a recalled mammogram (supplemental eTable 1), particularly for true-positive recalls with asymmetry of masses and microcalcifications (supplemental eTable 2 and supplemental eFigure 3).

Table 2. Association Between Nonmammographic and Mammographic Features and FP and TP Recalls

	Women With FP Recalls vs Women Who Were Not Recalled			Women With TP Recalls vs Women Who Were Not Recalled		
	FP Recall n (%)	Matched No Recall n (%)	OR (95% CI)[a]	TP Recall n (%)	Matched No Recall n (%)	OR (95% CI)[a]
Total, n	1,233	6,165		317	1,585	
Nonmammographic factors						
History of benign breast disease						
No	909 (75.6)	4,948 (81.5)	Ref	225 (73.3)	1,227 (78.7)	Ref
Yes	294 (24.4)	1,120 (18.5)	**1.44 (1.24–1.67)**	82 (26.7)	333 (21.3)	**1.36 (1.02–1.81)**
Family history of breast cancer						
No	1,021 (85.5)	5,182 (86.7)	Ref	236 (75.9)	1,340 (86.7)	Ref
Yes	173 (14.5)	793 (13.3)	1.11 (0.93–1.33)	75 (24.1)	206 (13.3)	**2.05 (1.52–2.76)**
Mammographic factors						
Dense area,[b] cm^2						
<9	157 (14.1)	1,033 (18.5)	Ref	65 (21.9)	398 (26.8)	Ref
9 to <20	180 (16.1)	1,075 (19.2)	1.12 (0.89–1.41)	60 (20.2)	330 (22.2)	1.13 (0.77–1.65)
≥20	780 (69.8)	3,477 (62.3)	**1.56 (1.28–1.89)**	172 (57.9)	757 (51.0)	**1.45 (1.05–2.02)**
Masses[b]						
0	409 (40.7)	2,953 (58.8)	Ref	66 (23.7)	786 (56.5)	Ref
≥1	595 (59.3)	2,067 (41.2)	**2.08 (1.81–2.39)**	212 (76.3)	604 (43.5)	**4.23 (3.14–5.72)**
Microcalcifications[b]						
0	856 (85.3)	4,503 (89.7)	Ref	164 (59.0)	1,190 (85.6)	Ref
≥1	148 (14.7)	517 (10.3)	**1.52 (1.25–1.86)**	114 (41.0)	200 (14.4)	**4.21 (3.15–5.63)**

Nonmammographic and mammographic features associated with mammography screening recalls based on Table 1 ($P<.1$).
Column totals may not equal the total number of subjects because of missing values.
Significant associations are highlighted in bold ($P<.05$).
Abbreviations: FR, false-positive; OR, odds ratio; TP, true-positive.
[a]Conditional logistic regression models in age-matched strata.
[b]Mammographic features from the recalled breast.

Breast Cancer Risk Models

Women with high risk scores—measured using Tyrer-Cuzick 10-year, KARMA 2-year, and polygenic risk models—were more likely to have a recalled mammogram (Table 1), including both false-positive and true-positive recalls (Table 3). Further analyses restricted to women who were recalled found that women with high risk scores were more likely to have a true-positive than a false-positive recall (Figure 2).

Discussion

In this large, population-based screening cohort, we identified several breast cancer risk factors and risk models that were associated with a higher risk of having a mammography recall. Further dividing recalls into false-positive and true-positive found that, among these factors, age was negatively associated with false-positive recall rates and positively associated with true-positive recall rates. Breast density and having a history of benign breast disease were equally associated with false-positive and true-positive recalls. Moreover, having a family history of breast cancer, masses, microcalcifications, and increased risk of breast cancer, measured using Tyrer-Cuzick, KARMA, and polygenic risk models, were associated with an increased risk of having a true-positive rather than a false-positive recall.

We found that 2.83% of women were recalled with false-positive recalls at the screening taken at enrollment in KARMA. In Sweden, all women aged 40 to 74 years are invited to attend breast cancer screening biennially (every 18–24 months) by mammography. Therefore, even if the rate of false-positive recalls is low in a single screening, a woman's lifetime risk of having a false-positive recall could be high. In the United States, half of women aged 40 to 69 years will have at least one false-positive mammogram after 10 screenings.[23] This large number of false-positive recalls, together with the fact that false-positive recalls can cause psychological burden,[3] highlights the importance of reducing false-positive recalls.

We showed that age was strongly associated with false-positive recalls in mammography screening. Among women

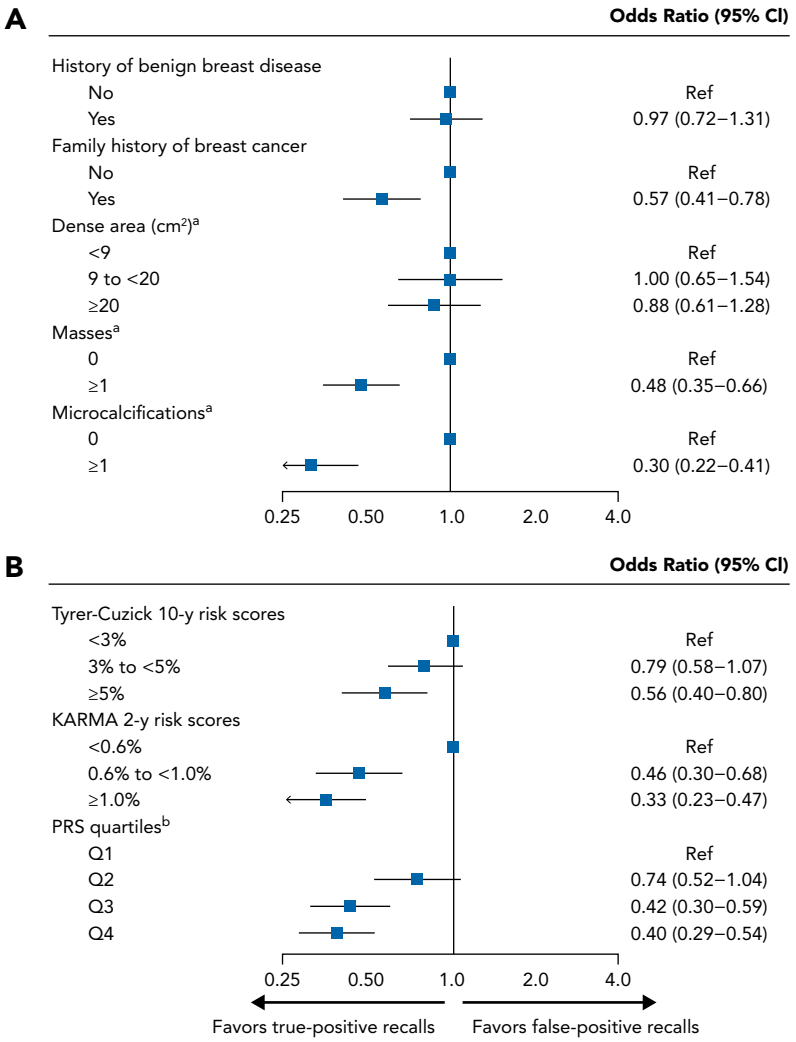


Figure 2. Association between **(A)** nonmammographic and mammographic features and **(B)** breast cancer risk models and false-positive recalls (compared with true-positive recalls), adjusting for age at mammography.
Abbreviations: KARMA, Karolinska Mammography Project for Risk Prediction of Breast Cancer; OR, odds ratio; PRS, polygenic risk score; Q, quartile.
[a]Mammographic features of the breast.
[b]Estimates from logistic regression models were adjusted for genotyping method and age at KARMA baseline. Analyses were conducted only in women with available genotype information. PRS quartile cutoffs were defined based on PRS distribution in women who were not recalled. False-/true-positive recalls were defined as women who ever received a false-/true-positive recall by 2015.

aged 40 to 49 years, there were, on average, 10 false-positive recalls per true-positive recall (tumor detected), which was almost 4 times higher than the observed rate among older women (aged >60 years). In part, this can be explained by younger women tending to have denser breasts than older women. In our study population, the mean dense area among women aged 40 to 49 years was 42.6 cm^2, which is higher than the 21.4 cm^2 observed among women aged >60 years. This finding suggests that young women generally may experience a higher harm/benefit ratio when attending mammography screening.[12] This finding has 2 important implications. First, for young women (aged 40–49 years) invited to screening, tailored interventions[24] to promote the benefits of screening while providing information about the likelihood of receiving a false-positive recall by age could be useful, because this experience may shape women's perceptions of mammography screening and therefore their future adherence. Second, given that there is still debate regarding whether to start mammography screening at 40 or 50 years of age, these concerns are important for policymakers to consider together with other factors to minimize the harm/benefit ratio and select the right women for mammography screening.

Mammographic features are the main basis for radiologists to determine screening results. In line with previous studies, we found that higher mammographic density is associated with false-positive recalls in mammography

Table 3. Association Between Breast Cancer Risk Models and FP and TP Recalls

	Women With FP Recalls vs Women Who Were Not Recalled			Women With TP Recalls vs Women Who Were Not Recalled		
	FP Recall n (%)	Matched No Recall n (%)	OR (95% CI)[a]	TP Recall n (%)	Matched No Recall n (%)	OR (95% CI)[a]
Total, n	1,233	6,165		317	1,585	
Tyrer-Cuzick 10-y risk scores						
<3%	670 (54.3)	3,490 (56.6)	Ref	112 (35.3)	695 (43.8)	Ref
3% to <5%	391 (31.7)	1,942 (31.5)	1.07 (0.92–1.25)	126 (39.7)	647 (40.8)	1.30 (0.96–1.74)
≥5%	172 (13.9)	733 (11.9)	**1.26 (1.03–1.54)**	79 (24.9)	243 (15.3)	**2.15 (1.53–3.02)**
KARMA 2-y risk scores						
<0.6%	752 (72.3)	4,275 (82.2)	Ref	101 (42.3)	880 (73.6)	Ref
0.6% to <1.0%	129 (12.4)	422 (8.1)	**1.83 (1.48–2.28)**	48 (20.1)	160 (13.4)	**2.75 (1.86–4.08)**
≥1.0%	159 (15.3)	503 (9.7)	**1.90 (1.56–2.33)**	90 (37.7)	155 (13.0)	**5.35 (3.78–7.57)**
PRS-quartiles[b]						
Q1	240 (21.2)	2,136 (25.0)	Ref	84 (10.2)	2,136 (25.0)	Ref
Q2	252 (22.2)	2,135 (25.0)	1.06 (0.88–1.28)	134 (16.3)	2,135 (25.0)	**1.58 (1.18–2.10)**
Q3	264 (23.3)	2,136 (25.0)	1.10 (0.91–1.33)	224 (27.3)	2,136 (25.0)	**2.77 (2.12–3.60)**
Q4	377 (33.3)	2,135 (25.0)	**1.59 (1.33–1.89)**	379 (46.2)	2,135 (25.0)	**4.49 (3.50–5.77)**

Breast cancer risk models associated with mammography screening recalls are based on Table 1 ($P<.1$).
Column totals may not equal the total number of subjects because of missing values.
Significant associations are highlighted in bold ($P<.05$).
Abbreviations: FP, false-positive; KARMA, Karolinska Mammography Project for Risk Prediction of Breast Cancer; OR, odds ratio; PRS, polygenic risk score; Q, quartile; TP, true-positive.
[a]Conditional logistic regression models in age-matched strata.
[b]Estimates from logistic regression models were adjusted for genotyping method and age at KARMA baseline. Analyses were conducted only in women with available genotype information. PRS quartile cutoffs were defined based on PRS distribution in women who were not recalled. False-/true-positive recalls were defined as women who ever received a false-/true-positive recall by 2015.

screening, suggesting that dense breasts may mask the appearance of the tumor, making it difficult to determine the screening result.[11,12] Our research question, however, was whether these factors were more strongly associated with false-positive than with true-positive recalls. Our findings suggest that high density is equally associated with both false-positive and true-positive recalls. Furthermore, our study showed that although microcalcifications and masses (identified through iCAD software) were positively associated with having a false-positive recall, they were actually more strongly associated with having a true-positive recall. The FDA-approved iCAD software is reluctantly used at some clinics, due to a fear of increasing false-positive recalls.[25,26] Our findings support the use of iCAD software to better detect breast tumors.

To the best of our knowledge, this is the first study to investigate and show that breast cancer risk models—Tyrer-Cuzick, KARMA, and polygenic risk models—are positively associated with having a false-positive recall. These results are novel but not surprising, given that these risk models directly or indirectly take mammographic features into account. Specifically, the KARMA model[27] directly incorporates breast density, masses, microcalcifications, and their asymmetries; the Tyrer-Cuzick model[21] incorporates hormone-related factors, which can affect mammographic features[28,29]; and the polygenic risk model positively correlates with not only mammographic density but also microcalcifications and masses.[30] However, when restricting our analyses to women who were recalled, we found that model estimated risks were associated with higher risk of having true-positive than false-positive recalls. Therefore, incorporating breast cancer risk models into mammography screening may help to identify true tumors rather than false-positive recalls. This finding is important because risk-based screening may be a reality in the near future.

We found that all investigated factors (except for age) either were not associated with false-positive recalls or were more closely associated with true-positive recalls. Therefore, these factors cannot be used to develop target interventions to diminish false-positive recalls. However, this does not mean that false-positive recalls cannot be minimized. For example, the recall rate in the United States is double that in Europe, even though cancer detection rates are similar,[23] indicating an unnecessary burden for women without breast cancer being recalled for further testing after screening. Furthermore, there are several novel

ways that may help to decrease false-positive recalls. First, an artificial intelligence support system might help radiologists to improve both the specificity and the sensitivity of mammography screening.[31] Second, other modalities, such as digital breast tomosynthesis and contrast-enhanced spectral mammography, have also been shown to reduce the number of false-positive recalls.[32,33]

Our study has several strengths. Our large sample size, together with detailed information collected in registers and a questionnaire, allows us to take a large number of breast cancer risk factors into consideration. We had full information on each woman's mammography screening history dated back to 1989, which guaranteed an accurate definition of screening results. Furthermore, we have measured breast density, masses, and microcalcifications using an automated method, thus strengthening the reproducibility and comparability of these mammographic features. Despite this, our study was limited to women participating in KARMA,[14] who are generally more highly educated and likely to have a family history of breast cancer than the general Swedish female population.[14] The generalizability of our results to other countries with different mammography screening strategies is limited, as with all studies investigating false-positive mammography recalls.[34]

Conclusions

Our study provides a better understanding of false-positive mammography recalls by comparison with both women who were not recalled and women who received true-positive recalls. Although several risk factors and risk models were associated with having a false-positive recall, they were equally or more strongly associated with having a true-positive recall. Our findings indicate that none of the studied breast cancer risk factors can be used to develop target interventions to reduce false-positive mammography recalls.

Submitted May 28, 2022; final revision received September 18, 2022; accepted for publication September 27, 2022.

Author contributions: *Conceptualization:* Mao, He, Czene. *Data curation:* Mao. *Formal analysis:* Mao, He. *Funding acquisition:* Mao, He, Humphreys, Czene. *Investigation:* All authors. *Methodology:* Mao, He, Humphreys, Czene. *Project administration:* He, Eriksson. *Resources:* Hall, Czene. *Software:* Mao, He, Humphreys, Eriksson. *Supervision:* He, Czene. *Validation:* Mao. *Visualization:* Mao. *Writing—original draft:* Mao, He, Czene. *Writing—review & editing:* All authors.

Disclosures: The authors have disclosed that they have not received any financial consideration from any person or organization to support the preparation, analysis, results, or discussion of this article.

Funding: Research reported in this publication was supported by Vetenskapsrådet under award numbers 2018-02547 (K. Czene) and 2020-01302 (K. Humphreys); Cancerfonden under award numbers 19 0266 (K. Czene) and 2020-0714 (K. Humphreys); FORTE under award number 2018-00877 (W. He); Stockholm County Council under award number 20200102 (K. Czene); China Scholarship Council under award number 20180621002 (X. Mao); and the Hundred Talents Program at Zhejiang University (W. He).

Correspondence: Wei He, PhD, Chronic Disease Research Institute, Department of Nutrition and Food Hygiene, the Children's Hospital, and National Clinical Research Center for Child Health, School of Public Health, School of Medicine, Zhejiang University, 866 Yu-hang-tang Road, Hangzhou, Zhejiang, 310058 China. Email: wei.he@ki.se

References

1. Mandelblatt JS, Stout NK, Schechter CB, et al. Collaborative modeling of the benefits and harms associated with different U.S. breast cancer screening strategies. Ann Intern Med 2016;164:215–225.
2. Yen AM, Tsau HS, Fann JC, et al. Population-based breast cancer screening with risk-based and universal mammography screening compared with clinical breast examination: a propensity score analysis of 1,429,890 Taiwanese women. JAMA Oncol 2016;2:915–921.
3. Weil JG, Hawker JI. Positive findings of mammography may lead to suicide. BMJ 1997;314:754–755.
4. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med 2007;146:502–510.
5. Canelo-Aybar C, Ferreira DS, Ballesteros M, et al. Benefits and harms of breast cancer mammography screening for women at average risk of breast cancer: a systematic review for the European Commission Initiative on Breast Cancer. J Med Screen 2021;28:389–404.
6. Tosteson AN, Fryback DG, Hammond CS, et al. Consequences of false-positive screening mammograms. JAMA Intern Med 2014;174:954–961.
7. Román R, Sala M, De La Vega M, et al. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. Breast Cancer Res Treat 2011;130:543–552.
8. McCann J, Stockton D, Godward S. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. Breast Cancer Res 2002;4:R11.
9. Lind H, Svane G, Kemetli L, et al. Breast cancer screening program in Stockholm County, Sweden - aspects of organization and quality assurance. Breast Care (Basel) 2010;5:353–357.
10. Elmore JG, Barton MB, Moceri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 1998;338:1089–1096.
11. Hubbard RA, Kerlikowske K, Flowers CI, et al. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med 2011;155:481–492.
12. Nelson HD, O'Meara ES, Kerlikowske K, et al. Factors associated with rates of false-positive and false-negative results from digital mammography screening: an analysis of registry data. Ann Intern Med 2016;164:226–235.
13. Ho PJ, Bok CM, Ishak HMM, et al. Factors associated with false-positive mammography at first screen in an Asian population. PLoS One 2019;14:e0213615.
14. Gabrielson M, Eriksson M, Hammarström M, et al. Cohort profile: the Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). Int J Epidemiol 2017;46:1740–1741g.
15. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, et al. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur J Epidemiol 2009;24:659–667.
16. Holm J, Humphreys K, Li J, et al. Risk factors and tumor characteristics of interval cancers by mammographic density. J Clin Oncol 2015;33:1030–1037.
17. Socialstyrelsen [National Board of Health and Welfare]. Screening för bröstcancer: rekommendation och bedömningsunderlag [Screening for breast cancer: recommendation and assessment basis]. Accessed January 15, 2022. Available at: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/nationella-screeningprogram/2014-2-32.pdf
18. Olsson S, Andersson I, Karlberg I, et al. Implementation of service screening with mammography in Sweden: from pilot study to nationwide programme. J Med Screen 2000;7:14–18.
19. Li J, Szekely L, Eriksson L, et al. High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. Breast Cancer Res 2012;14:R114.

20. Eriksson M, Czene K, Pawitan Y, et al. A clinical model for identifying the short-term risk of breast cancer. Breast Cancer Res 2017;19:29.
21. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med 2004;23:1111–1130.
22. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 2019;104:21–34.
23. Fletcher SW, Elmore JG. False-positive mammograms—can the USA learn from Europe? Lancet 2005;365:7–8.
24. Sabatino SA, Lawrence B, Elder R, et al. Effectiveness of interventions to increase screening for breast, cervical, and colorectal cancers: nine updated systematic reviews for the guide to community preventive services. Am J Prev Med 2012;43:97–118.
25. Khoo LA, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. Radiology 2005;237:444–449.
26. Cole EB, Zhang Z, Marques HS, et al. Impact of computer-aided detection systems on radiologist accuracy with digital mammography. AJR Am J Roentgenol 2014;203:909–916.
27. Michailidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551:92–94.
28. Boyd NF, Rommens JM, Vogt K, et al. Mammographic breast density as an intermediate phenotype for breast cancer. Lancet Oncol 2005;6:798–808.
29. Azam S, Eriksson M, Sjölander A, et al. Predictors of mammographic microcalcifications. Int J Cancer 2021;148:1132–1143.
30. Holowko N, Eriksson M, Kuja-Halkola R, et al. Heritability of mammographic breast density, density change, microcalcifications, and masses. Cancer Res 2020;80:1590–1600.
31. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology 2019;290:305–314.
32. Mann RM, Hooley R, Barr RG, et al. Novel approaches to screening for breast cancer. Radiology 2020;297:266–285.
33. Chong A, Weinstein SP, McDonald ES, et al. Digital breast tomosynthesis: concepts and clinical practice. Radiology 2019;292:1–14.
34. Elmore JG, Nakano CY, Koepsell TD, et al. International variation in screening mammography interpretations in community-based programs. J Natl Cancer Inst 2003;95:1384–1393.

See JNCCN.org for supplemental online content.

JNCCN

National Comprehensive Cancer Network®

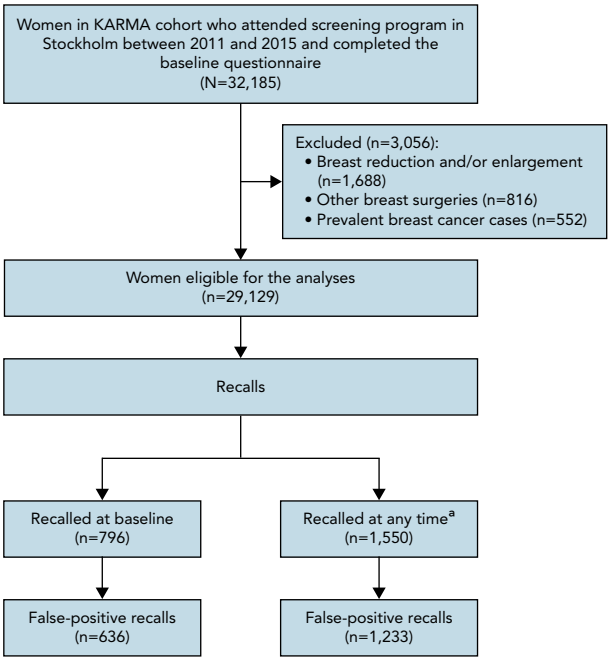
JOURNAL OF THE NATIONAL COMPREHENSIVE CANCER NETWORK

Supplemental online content for:

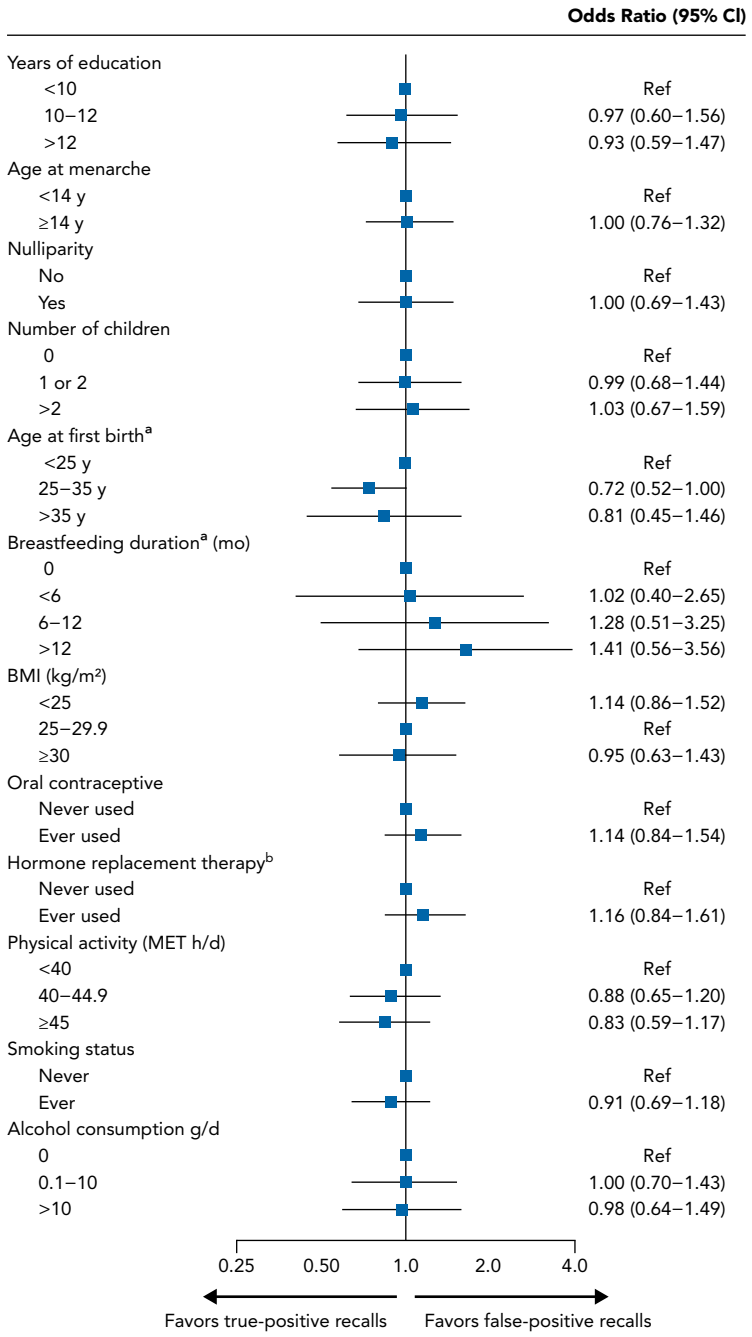
Factors Associated With False-Positive Recalls in Mammography Screening

Xinhe Mao, MD, MSc; Wei He, PhD; Keith Humphreys, PhD; Mikael Eriksson, PhD; Natalie Holowko, PhD; Fredrik Strand, MD, PhD; Per Hall, MD, PhD; and Kamila Czene, PhD

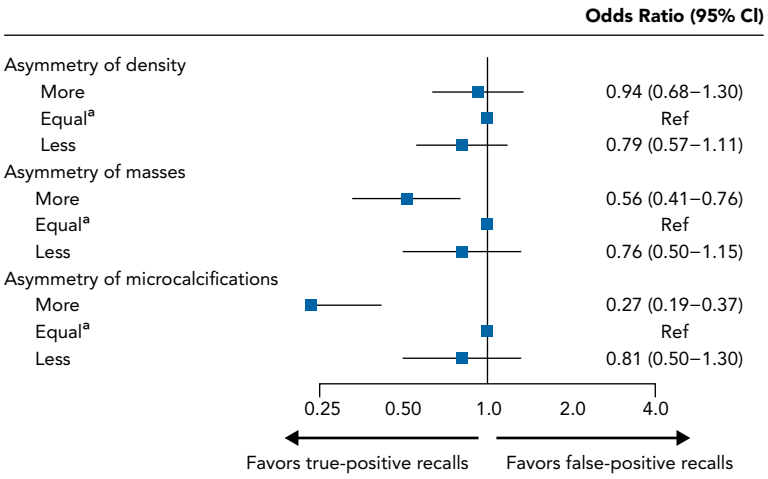
J Natl Compr Canc Netw 2023;21(2):143–152.e4



eFigure 1. Study populations used for analyses.
Abbreviation: KARMA, Karolinska Mammography Project for Risk Prediction of Breast Cancer.
[a]Women who received their first mammography recall at or after KARMA baseline.



eFigure 2. Association between breast cancer risk factors and false-positive recalls (vs true-positive recalls).
Breast cancer risk factors not associated with mammography screening recalls based on Table 1 ($P \geq .1$) are included.
Abbreviations: BMI, body mass index; MET, metabolic equivalents of task.
[a]Among parous women.
[b]Among postmenopausal women.



eFigure 3. Association between bilateral symmetry of mammographic features (compared with contralateral side) and false-positive recalls (compared with true-positive recalls), adjusting for age at mammogram. Asymmetry of mammographic features were defined as the recalled side dense area, number of masses, and number of microcalcification clusters minus those on the contralateral side.
[a]Equal was defined as within 6 cm^2 of dense area, same number of masses, and same number of microcalcification clusters, respectively.

eTable 1. Association Between Bilateral Symmetry of Mammographic Features[a] and Mammography Screening Recall Rates

	Number of Screenings	Recalls (Crude rate per 1,000 screenings)		*P* Value[b]
		Yes	No	
Dense area on the recalled compared with contralateral side[c]				.002
More	6,377	211 (33.1)	6,166 (966.9)	
Equal[d]	15,412	382 (24.8)	15,030 (975.2)	
Less	6,384	184 (28.8)	6,200 (971.2)	
Masses on the recalled compared with contralateral side[c]				<.001
More	5,841	264 (45.2)	5,577 (954.8)	
Equal[d]	16,726	404 (24.2)	16,322 (975.8)	
Less	5,606	109 (19.4)	5,497 (980.6)	
Microcalcifications on the recalled compared with contralateral side[c]				<.001
More	1,887	97 (51.4)	1,790 (948.6)	
Equal[d]	24,573	625 (25.4)	23,948 (974.6)	
Less	1,713	55 (32.1)	1,658 (967.9)	

Column totals may not equal to total number of subjects due to missing values.
[a]Compared with contralateral side.
[b]*P* values from chi-square tests.
[c]Asymmetries of mammographic features were defined as the recalled side dense area, number of masses, and number of microcalcification clusters minus those on the contralateral side.
[d]Equal was defined as within 6 cm^2 of dense area, same number of masses, and same number of microcalcification clusters, respectively.

eTable 2. Association Between Bilateral Symmetry of Mammographic Features[a] and FP and TP Recalls

	Women With FP Recalls vs Women Who Were Not Recalled			Women With TP Recalls vs Women Who Were Not Recalled		
	FP Recall n (%)	Matched No Recall n (%)	OR (95% CI)[b]	TP Recall n (%)	Matched No Recall n (%)	OR (95% CI)[b]
Total, n	1,233	6,165		317	1,585	
Dense area on the recalled compared with contralateral side[c]						
More	328 (29.4)	1,343 (24.0)	**1.35 (1.16–1.57)**	77 (25.9)	350 (23.6)	1.24 (0.91–1.67)
Equal[d]	519 (46.5)	2,856 (51.1)	Ref	145 (48.8)	814 (54.8)	Ref
Less	270 (24.2)	1,386 (24.8)	1.08 (0.92–1.27)	75 (25.3)	321 (21.6)	1.32 (0.97–1.80)
Masses on the recalled compared with contralateral side[c]						
More	409 (40.7)	1,323 (26.4)	**1.78 (1.53–2.07)**	153 (55.0)	368 (26.5)	**3.43 (2.53–4.66)**
Equal[d]	417 (41.5)	2,403 (47.9)	Ref	80 (28.8)	647 (46.5)	Ref
Less	178 (17.7)	1,294 (25.8)	**0.80 (0.66–0.96)**	45 (16.2)	375 (27.0)	0.98 (0.67–1.45)
Microcalcifications on the recalled compared with contralateral side[c]						
More	115 (11.5)	396 (7.9)	**1.57 (1.26–1.96)**	98 (35.3)	153 (11.0)	**4.94 (3.60–6.79)**
Equal[d]	800 (79.7)	4,268 (85.0)	Ref	151 (54.3)	1,112 (80.0)	Ref
Less	89 (8.9)	356 (7.1)	**1.36 (1.06–1.74)**	29 (10.4)	125 (9.0)	**1.87 (1.20–2.92)**

Asymmetric mammographic features associated with mammography screening recalls based on supplemental eTable 1 ($P<.1$).
Column totals may not equal to total number of subjects due to missing values.
Significant associations are highlighted in bold ($P<.05$).
Abbreviations: FP, false-positive; OR, odds ratios; TP, true-positive.
[a]Compared with contralateral side.
[b]Conditional logistic regression models in age-matched strata.
[c]Asymmetries of mammographic features were defined as the recalled side dense area, number of masses, and number of microcalcification clusters minus those on the contralateral side.
[d]Equal was defined as within 6 cm^2 of dense area, same number of masses, and same number of microcalcification clusters, respectively.