

# Impact of Artificial Intelligence System and Volumetric Density on Risk Prediction of Interval, Screen-Detected, and Advanced Breast Cancer

Celine M. Vachon, PhD<sup>1</sup>; Christopher G. Scott, MS<sup>2</sup>; Aaron D. Norman, MPH<sup>1</sup>; Sadia A. Khanani, MD<sup>3</sup>; Matthew R. Jensen, BS<sup>2</sup>; Carrie B. Hruska, PhD<sup>3</sup>; Kathleen R. Brandt, MD<sup>3</sup>; Stacey J. Winham, PhD<sup>4</sup>; and Karla Kerlikowske, MD<sup>5</sup>

**PURPOSE** Artificial intelligence (AI) algorithms improve breast cancer detection on mammography, but their contribution to long-term risk prediction for advanced and interval cancers is unknown.

**METHODS** We identified 2,412 women with invasive breast cancer and 4,995 controls matched on age, race, and date of mammogram, from two US mammography cohorts, who had two-dimensional full-field digital mammograms performed 2-5.5 years before cancer diagnosis. We assessed Breast Imaging Reporting and Data System density, an AI malignancy score (1-10), and volumetric density measures. We used conditional logistic regression to estimate odds ratios (ORs), 95% CIs, adjusted for age and BMI, and C-statistics (AUC) to describe the association of AI score with invasive cancer and its contribution to models with breast density measures. Likelihood ratio tests (LRTs) and bootstrapping methods were used to compare model performance.

**RESULTS** On mammograms between 2-5.5 years prior to cancer, a one unit increase in AI score was associated with 20% greater odds of invasive breast cancer (OR, 1.20; 95% CI, 1.17 to 1.22; AUC, 0.63; 95% CI, 0.62 to 0.64) and was similarly predictive of interval (OR, 1.20; 95% CI, 1.13 to 1.27; AUC, 0.63) and advanced cancers (OR, 1.23; 95% CI, 1.16 to 1.31; AUC, 0.64) and in dense (OR, 1.18; 95% CI, 1.15 to 1.22; AUC, 0.66) breasts. AI score improved prediction of all cancer types in models with density measures ( $P_{LRT}$  values < .001); discrimination improved for advanced cancer (ie, AUC for dense volume increased from 0.624 to 0.679,  $\Delta$  AUC 0.065,  $P = .01$ ) but did not reach statistical significance for interval cancer.

**CONCLUSION** AI imaging algorithms coupled with breast density independently contribute to long-term risk prediction of invasive breast cancers, in particular, advanced cancer.

**J Clin Oncol 41:3172-3183. © 2023 by American Society of Clinical Oncology**

## INTRODUCTION

Multiple artificial intelligence (AI) algorithms based on deep learning have been developed, which improve accuracy of breast cancer detection on mammography.<sup>1-4</sup> More recently, these algorithms have been considered for their use in near-term risk prediction and preliminary data suggest improved discrimination of breast cancer compared with conventional clinical risk models.<sup>1,2</sup>

There are limited data, however, for the performance of these AI algorithms in predicting long-term risk of invasive cancer and risk of advanced and interval cancers (defined as screening failures), which would most benefit from early detection and intervention. Interval cancers are known to be increased among women with dense breasts,<sup>5</sup> because of decreased mammogram sensitivity.<sup>6,7</sup> Wanders et al<sup>8</sup> recently validated an AI cancer detection system integrating area-based breast density to improve discrimination of interval cancer

within 2 years of a negative mammogram. They did not have information on screen-detected cancers, or consider advanced cancers, or long-term risk. Understanding long-term risk prediction models that incorporate AI for advanced cancer will provide insight into their ability to guide screening strategies to reduce breast cancer morbidity and mortality.<sup>9,10</sup>

In this article, we extend prior work to evaluate the contribution of an AI cancer detection system on the basis of deep convolutional neural networks combined with volumetric density measures to interval, screen-detected, advanced, and nonadvanced cancer risk. We hypothesize that the AI detection score, which reflects suspicious mammographic findings, will improve long-term risk prediction of invasive cancer above volumetric density measures and for advanced and interval breast cancers. We focus on assessing mammograms taken at least 2 years but not more than 5.5 years before the cancer to evaluate longer-term risk prediction.

## ASSOCIATED CONTENT

### Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on February 24, 2023 and published at [ascopubs.org/journal/jco](https://ascopubs.org/journal/jco) on April 27, 2023; DOI <https://doi.org/10.1200/JCO.22.01153>

## CONTEXT

### Key Objective

Does an artificial intelligence (AI) algorithm designed for improved breast cancer detection on mammography contribute to long-term breast cancer risk prediction, in particular, for interval and advanced cancers, and in combination with breast density?

### Knowledge Generated

In a nested case-control study of 2,412 invasive breast cancers and 4,995 matched controls, an AI score assessed on mammograms 2-5.5 years before cancer was associated with invasive breast cancer and similarly predictive of interval and advanced cancers and in dense and nondense breasts. Furthermore, the score improved prediction of all cancer types in models with density measures. AI imaging algorithms coupled with breast density independently contribute to long-term risk prediction of invasive breast cancers, including advanced cancers.

### Relevance (P.H. Frankel)

This effort reports a computer-aided/AI mammogram-based risk model to improve risk estimates excluding breast cancer diagnosis within 2 years from a negative mammogram. This adds further evidence to support computer-aided imaging assessments and will help motivate prospective studies to disentangle improved earlier diagnosis from an improved risk model for subsequent cancer. Such prospective studies will also provide estimates that can more easily guide clinical practice.\*

\*Relevance section written by JCO Associate Editor Paul H. Frankel, PhD.

## METHODS

### Study Sample

Study participants were from two retrospective case-control studies nested within large breast imaging cohorts from US-based breast screening practices (Fig 1). The San Francisco Mammography Registry (SFMR) participates in the National Cancer Institute–funded Breast Cancer Surveillance Consortium (BCSC).<sup>11</sup> The SFMR obtains annual institutional review board approval and passive permission for data collection and participant enrollment and data linkages for research purposes. The Mayo Clinic institutional review board approved a waiver of informed consent and Health Insurance Portability and Accountability Act authorization for participants from the breast screening practice who did not refuse permission to use their medical records for research (according to Minnesota Research Authorization).

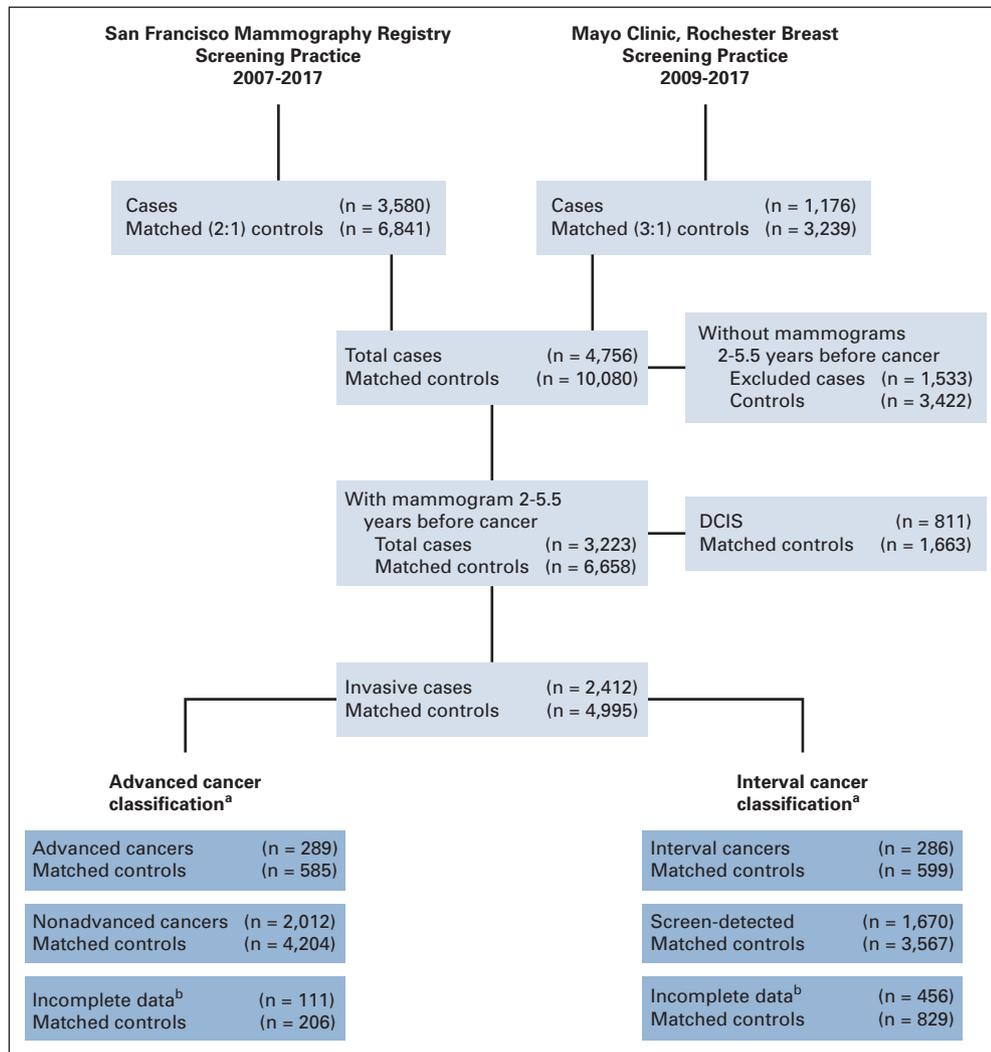
The SFMR obtained for-processing digital screening examinations from Hologic Selenia machines at four facilities since 2006, which served as the underlying SFMR imaging cohort. Annual linkage to the California Cancer Registry identified 3,580 incident breast cancers reported from 2007 through 2017. Up to two control participants ( $n = 6,841$ ) without previous breast cancer or breast implants were selected from the SFMR imaging cohort and matched to each case participant by age within 5 years, race, date of screening examination within 1 year, mammography machine, and facility.

For the Mayo Clinic screening cohort, digital images from Hologic Selenia machines were collected from women in the tristate region of Minnesota, Iowa, and Wisconsin

at the Rochester, Minnesota facility from 2009 through 2017. Annual linkage to the Mayo Clinic tumor registry identified cases of incident breast cancer reported through December 2017 ( $n = 1,176$ ). Up to three control participants ( $n = 3,239$ ) without previous breast cancer or breast implants were selected from the Mayo Clinic cohort and matched on the same criteria as above.

Primary analyses were restricted to invasive cases only with mammograms between 2 and 5.5 years before diagnosis or last follow-up, resulting in 1,700 cases and 3,109 matched controls from SFMR and 712 invasive cancers and 1,886 matched controls from Mayo (Fig 1). For participants with multiple images within the time window of 2-5.5 years before cancer, only one mammogram (closest to 5 years before) was used for all cases and their matched controls from both studies.

Interval cancer was defined as invasive breast cancer occurring within 12 months of a negative mammography result (Breast Imaging Reporting and Data System [BI-RADS] 1 or 2). Screen-detected cancer was defined as invasive cancer occurring within 12 months of a positive screening mammography result (BI-RADS 0) with a final BI-RADS of 3, 4, or 5. Advanced cancer was defined using American Joint Committee on Cancer (AJCC) prognostic pathologic stage II or higher since it most accurately predicts breast cancer death.<sup>10</sup> All advanced cancers were included regardless of mammography BI-RADS assessment. Women without mammograms within the appropriate time window (for defining interval and screen-detected cancers) or tumor characteristics (for defining advanced cancers) were classified as missing (Fig 1).



**FIG 1.** Study schema. <sup>a</sup>Invasive breast cancers were classified as advanced cancer and interval cancer, two primary end points for the study. These are not mutually exclusive categorizations. <sup>b</sup>Unable to classify because of missing clinical data (as seen in the Data Supplement [Supplemental Table 1]). DCIS, ductal carcinoma in situ.

### AI Cancer Detection System

A validated AI detection system (Transpara, version 1.7.0; ScreenPoint Medical) was used to provide a score from 1-10, representing an increasing probability of malignancy at the time of breast screening.<sup>3,4,12,13</sup> The system is based on deep convolutional neural networks and automatically detects signs of malignancy on digital mammography and digital breast tomosynthesis.<sup>4</sup> All four mammogram views were deidentified and processed in a blinded fashion without knowledge of clinical information to arrive at the malignancy score.

### BI-RADS Breast Density and BI-RADS Assessment

Clinical BI-RADS density was classified into four categories as part of routine clinical practice: (a) almost entirely fatty, (b) scattered fibroglandular densities, (c) heterogeneously

dense, and (d) extremely dense. Volumetric density was assessed in a blinded fashion using Volpara, version 1.5.4, a fully automated method that uses the measured breast thickness and x-ray attenuations in the for-processing image to create estimates of dense volume (DV) and volumetric percent density (VPD) measures for each woman, incorporating all four mammogram views,<sup>14</sup> as performed in the clinical setting.

In addition, clinical BI-RADS assessment that describes mammogram interpretation (0 = incomplete, 1 = negative, 2 = benign findings, 3 = probably benign, 4 = suspicious abnormality, and 5 = highly suspicious of malignancy) was available from the clinical review of mammograms from the Mayo Clinic cohort only.

## Statistical Analyses

We used an existing study sample designed to examine the association of automated density measures with breast cancer risk.<sup>14,15</sup> Descriptive characteristics were compared by case-control status, for both studies combined and separately; Kruskal-Wallis and chi-squared tests were used to examine differences by case status.

Because of the matched study design, conditional logistic regression was used to examine primary associations of AI score, breast density measures (BI-RADS density, and log-transformed VPD%, DV), and breast cancer risk, including invasive breast cancer, advanced/nonadvanced cancer, and interval/screen-detected cancer and secondarily, ductal carcinoma in situ (DCIS). Odds ratios (ORs) and 95% CIs were estimated. C-statistics (and 95% CIs) were also calculated as a measure of discriminatory accuracy or AUC. Models assumed a linear trend for Transpara, VPD%, and DV and an ordinal trend for BI-RADS density. The C-statistics accounted for the matched data by taking the ratio of the total number of concordant pairs and the total number of pairs, considering only pairs within matched sets.<sup>16</sup>

Subgroup analyses were performed to examine differences in AI score and breast cancer associations by age (<60 and 60+ years), dense breasts (BI-RADS a, b v c, d), BMI (<25, 25-30, and 30+), and race/ethnicity (Asian, Black, Hispanic, White, and Others). We also examined models with AI score and BI-RADS assessment, using similar approaches, among the Mayo Clinic studies, where data were available. All models were adjusted for continuous age and BMI at the time of mammogram. Stratified analyses of BI-RADS density required breaking the matched data and were analyzed using unconditional logistic regression, adjusted for the matching factors.

We compared the performance of nested models (ie, addition of AI to breast density models) using the likelihood ratio test (LRT;  $P_{LRT}$ ) to evaluate improvement in model fit, as previously described.<sup>17</sup> We also evaluated improvement in discriminatory accuracy by comparing AUCs between models on the basis of 1,000 bootstrap samples. Each bootstrap sample was constructed by selecting n-matched sets of individuals chosen at random with replacement, with n equal to the number of cases in the specific analysis. Matched C-statistics were calculated as described above for each model using each bootstrap sample. For each model comparison, the difference in model C-statistics was calculated over all 1,000 samples; 95% confidence limits and normal distribution *P* values ( $P_{AUC}$ ) were calculated using the standard deviation of this difference as the SE estimate. We used the same bootstrapping approach to test differences in AUCs across subgroups ( $P_{AUC}$ ). SAS version 9.4 (Cary, NC) was used for analyses, and two-sided *P* values < .017 (0.05 ÷ 3) were considered statistically significant to account for multiple testing of invasive, interval, and advanced cancer.

## RESULTS

A total of 2,412 participants with invasive breast cancer and 4,995 matched control participants were eligible from the two screening cohorts (Table 1; Fig 1; and Data Supplement [Supplemental Table 1], online only). Of the invasive cases, 286 were interval cancers, 1,670 were screen-detected, and 456 were missing, most of whom (83%) did not have imaging within the required time points for classification; there were 289 advanced cancers, 2,012 nonadvanced cancers, and 111 missing an essential tumor characteristic required for AJCC prognostic stage classification (Data Supplement [Supplemental Table 2]). The median follow-up time was 4.1 years, and 47% of participants were older than age 60 years. BMI, breast density, and AI scores were higher among cases versus controls (Table 1). AI detection scores were weakly correlated with breast density measures, age, and BMI (Data Supplement [Supplemental Table 3]). Mayo participants were more likely to be older, be White, have a higher BMI, and have nondense breasts than SFMR participants (Data Supplement [Supplemental Table 1]).

### AI Malignancy Score and Breast Cancer Risk

Adjusted for age and BMI, AI score was associated with invasive breast cancer (OR, 1.20; 95% CI, 1.17 to 1.22 per unit increase; AUC, 0.63; 95% CI, 0.62 to 0.64). ORs and discriminatory accuracy were similar for interval and screen-detected cancers ( $P_{AUC} = .93$ ; Table 2), advanced and nonadvanced cancers ( $P_{AUC} = .67$ ; Table 2), and dense and nondense breasts ( $P_{AUC} = .19$ ; Fig 2). There was an increased cancer risk and higher discrimination associated with AI score among women age 60 years and older compared with those age younger than 60 years ( $\Delta AUC$ , 0.04; 95% CI, 0.01 to 0.07;  $P_{AUC} = .01$ ; Fig 2) but no statistically significant difference in discrimination by obesity (BMI > 30 relative to normal or underweight BMI < 25 [ $\Delta AUC$ , 0.03; 95% CI, 0.00 to 0.07;  $P_{AUC} = .05$ ]) or race (all  $P_{AUC} > .37$ ; Fig 2).

### Association of Breast Density Measures With Breast Cancer Risk

Adjusted for age and BMI, all breast density measures were associated with invasive breast cancer (Table 3). Unlike the AI score, there were stronger associations and discrimination of interval (v screen-detected) cancer for all density measures ( $P_{AUC}$  values < .02; Table 3). For advanced versus nonadvanced cancer, the ORs and AUCs were similar for all density measures ( $P_{AUC}$  values > .14; Table 3).

### Contribution of AI to Breast Density Measures

The inclusion of AI scores combined with either clinical or volumetric density measures improved prediction of all cancer types compared with models with density alone (Table 3; All  $P_{LRT} < .001$ ). For advanced cancers, inclusion of AI scores with DV also resulted in statistically significant improvement in discriminatory accuracy relative to models

**TABLE 1.** Demographics, Breast Density, and AI Score Distributions by Invasive Breast Cancer (case-control) Status

Characteristic	Case (n = 2,412)	Control (n = 4,995)	P
Age, years			.81
Mean (SD)	58.9 (11.7)	59.0 (11.5)	
Range	33-94	32-93	
BMI			.02
Mean (SD)	26.5 (5.8)	26.2 (6.0)	
Median	25.1	24.8	
Missing	39	68	
Race/ethnicity, No. (%)			.15
Asian	381 (15.8)	713 (14.3)	
Black	63 (2.6)	100 (2.0)	
Hispanic	52 (2.2)	103 (2.1)	
Native American	1 (0.0)	2 (0.0)	
Others/Mixed	89 (3.7)	152 (3.0)	
Unknown	3 (0.1)	9 (0.2)	
White	1,823 (75.6)	3,916 (78.4)	
Follow-up time, years, <sup>a</sup> No. (%)			.60
2-3	531 (22.0)	1,102 (22.1)	
3-4	612 (25.4)	1,319 (26.4)	
4-5.5	1,269 (52.6)	2,574 (51.5)	
BI-RADS density, No. (%)			<.0001
a	235 (10.0)	762 (15.6)	
b	823 (35.0)	1,956 (40.0)	
c	972 (41.3)	1,727 (35.3)	
d	321 (13.7)	450 (9.2)	
Missing	61	100	
Volpara volumetric density			
Mean VPD (SD)	10.4 (6.8)	9.3 (6.2)	<.0001
Mean DV (SD)	68.3 (38.3)	58.0 (31.2)	<.0001
AI score, No. (%)			<.0001
1	75 (3.1)	380 (7.6)	
2	81 (3.4)	292 (5.8)	
3	134 (5.6)	478 (9.6)	
4	170 (7.0)	568 (11.4)	
5	260 (10.8)	690 (13.8)	
6	246 (10.2)	535 (10.7)	
7	274 (11.4)	561 (11.2)	
8	298 (12.4)	542 (10.9)	
9	361 (15.0)	506 (10.1)	
10	513 (21.3)	443 (8.9)	
AI score, mean (SD)	6.5 (2.6)	5.2 (2.7)	
Advanced cancer, No. (%)			
Advanced	289 (12.6)	NA	NA
Nonadvanced	2,012 (87.4)	NA	

(continued on following page)

**TABLE 1.** Demographics, Breast Density, and AI Score Distributions by Invasive Breast Cancer (case-control) Status (continued)

Characteristic	Case (n = 2,412)	Control (n = 4,995)	P
Incomplete data <sup>b</sup>	111	NA	
Interval cancer, No. (%)			
Interval	286 (14.6)	NA	NA
Screen-detected	1,670 (85.4)	NA	
Incomplete data <sup>b</sup>	456	NA	

Abbreviations: AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System; DV, dense volume; SD, standard deviation; VPD, volumetric percent density.

<sup>a</sup>For cases, follow-up to cancer. For controls, follow-up from mammogram to matched date of case.

<sup>b</sup>Incomplete clinical data were available to classify cases (see the Data Supplement [Supplemental Table 1]).

with DV alone ( $\Delta$ AUC, 0.065; 95% CI, 0.015 to 0.114;  $P_{AUC} = .01$ ). Similar improvement was observed with addition of AI scores to VPD ( $\Delta$ AUC, 0.067; 95% CI, 0.018 to 0.116;  $P_{AUC} = .007$ ) but did not reach statistical significance with BI-RADS ( $\Delta$ AUC, 0.045; 95% CI,  $-0.006$  to 0.097;  $P_{AUC} = .09$ ; Table 3). Improvement in discrimination was also seen with AI scores added to density models for non-advanced cancers, for all three density measures (VPD, DV, and BI-RADS) compared with models with density alone ( $P_{AUC}$  values  $< .001$ ; Table 3).

As expected, all AUCs for screen-detected cancer improved with inclusion of AI scores for models with VPD, DV, and BI-RADS density ( $P_{AUC}$  values  $\leq .001$ ; Table 3). However, for interval cancers, the contribution of AI scores resulted in smaller gains in AUC, which did not reach statistical significance for models with any density measure (Table 3).

### Secondary Analyses

We examined the association of the AI score with 811 in situ cancers and 1,663 matched controls with mammograms 2-5.5 years before breast cancer (Fig 1). The associations of AI score and in situ breast cancer were like invasive cancer (OR, 1.16; 95% CI, 1.13 to 1.20; AUC, 0.62; 95% CI, 0.59 to 0.64) and showed similar improvement in risk prediction ( $P_{LRT} < .001$ ) and discrimination ( $P_{AUC}$  values  $< .01$ ) with the addition of AI to all density measures (Data Supplement [Supplemental Table 4]).

We also examined the contribution of AI score to the BI-RADS assessment and breast cancer association. As expected, BI-RADS assessment was associated with breast cancer

(Data Supplement [Supplemental Tables 5 and 6]), with increased odds of breast cancer among all categories relative to BI-RADS 1 or negative findings; the AUC was 0.566 (95% CI, 0.543 to 0.587). However, across all BI-RADS assessments, the mean AI score was higher among cases than controls (Data Supplement [Supplemental Figure 1]). The addition of the AI score resulted in attenuation of the risk estimates for BI-RADS assessment and a higher AUC (0.643; 95% CI, 0.621 to 0.664). Importantly, the AI score and breast cancer association did not change in models with (OR, 1.20; 95% CI, 1.16 to 1.24) or without BI-RADS assessment (OR, 1.19; 95% CI, 1.15 to 1.23).

In addition, we examined how Transpara performed in our study population closer to the cancer for comparison with prior studies. Of our invasive case-control sample, 1,010 cases and 2,140 controls also had mammograms within 2 years of the cancer (cases) or follow-up date (controls). We found statistically significant, but stronger association and higher discrimination of AI models (without density measures) within 2 years of cancer (OR, 1.30; 95% CI, 1.26 to 1.34; AUC, 0.70; 95% CI, 0.68 to 0.72) relative to AI models assessed on mammograms 2-5.5 years before the cancer (OR, 1.20; 95% CI, 1.17 to 1.24; AUC, 0.63; 95% CI, 0.61 to 0.65).

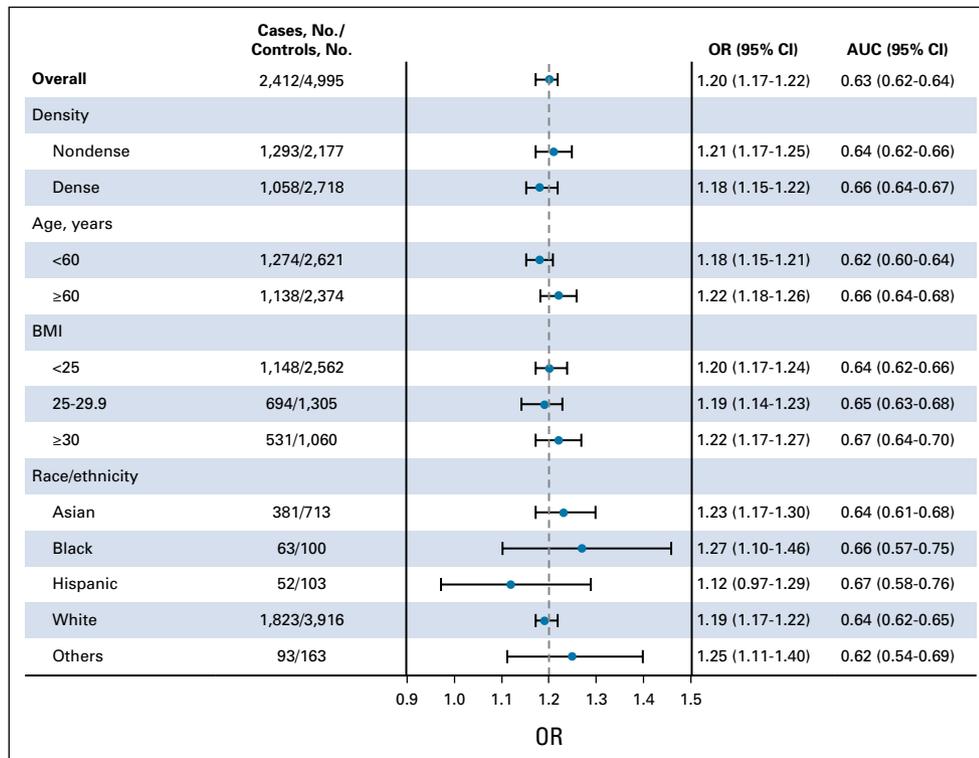
### DISCUSSION

The Transpara AI algorithm assessed on mammograms 2-5.5 years before breast cancer showed similar associations with invasive cancer and discriminatory accuracy

**TABLE 2.** Association of Transpara Artificial Intelligence Malignancy Score With Breast Cancer Risk

Breast Cancer Outcome	Cases, No./Controls, No.	OR (95% CI)	AUC (95% CI)
All invasive cancer	2,412/4,995	1.20 (1.17 to 1.22)	0.630 (0.617 to 0.644)
Screen-detected	1,670/3,567	1.21 (1.18 to 1.24)	0.641 (0.625 to 0.657)
Interval cancer	286/599	1.20 (1.13 to 1.27)	0.631 (0.592 to 0.670)
Advanced cancer	289/585	1.23 (1.16 to 1.31)	0.638 (0.599 to 0.677)
Nonadvanced cancer	2,012/4,204	1.19 (1.17 to 1.22)	0.630 (0.616 to 0.645)

Abbreviation: OR, odds ratio.



**FIG 2.** Associations (ORs, 95% CI) of Transpara AI malignancy score with breast cancer by age, density, BMI, and race/ethnicity. AI, artificial intelligence; OR, odds ratio.

in dense and nondense breasts and for interval versus screen-detected and advanced versus nonadvanced cancers. These findings imply that the AI score is robust for long-term risk prediction by severity of cancer and extent of density. Furthermore, the incorporation of AI score into models combined with breast density showed improved prediction for all invasive cancer types, with the highest discriminatory accuracy seen for advanced and interval cancers, suggesting that the combination will be important for prediction of these cancers with worse prognosis.

Previous studies have developed and validated AI algorithms, including Mirai and iCAD, that show high discrimination of breast cancer<sup>1,2,18-20</sup> (invasive cancer and DCIS combined) with AUCs above 0.75, surpassing that of established clinical risk models, such as Gail and Tyrer-Cuzick/IBIS.<sup>1,2</sup> iCAD was developed for a near-term 2-year risk prediction, whereas Mirai was developed for risk prediction up to 5 years. Although our AUCs for longer-term risk are lower than those reported by Mirai, it is difficult to directly compare for multiple reasons. First, we excluded mammograms within 2 years of the cancer, where the algorithms show optimal performance, whereas these previous studies using Mirai and iCAD did not. When we investigated the performance of Transpara only using mammograms within 2 years of the cancer, we show higher discrimination (AUC, 0.70), comparable with these other studies. Second, our case-control study design matched

closely on age, mammogram date, and race/ethnicity, all of which limit the ability to examine their contributions to AUC and direct comparison across other study designs. Finally, we evaluated invasive breast cancer that may benefit from targeted screening, whereas Mirai and iCAD algorithms include invasive cancer and DCIS. Ensuring accurate prediction of the cancers with worst prognosis, that is, advanced cancer defined as pathologic prognostic stage, should be a priority of AI risk models to allow early intervention or prevention. In fact, the TMIST clinical trial is using study-defined definition of advanced cancer as its primary end point<sup>21</sup> although their definition has been shown to be less predictive of breast cancer death than prognostic pathologic stage.<sup>9</sup> We found that the Transpara AI score assessed on mammograms 2-5.5 years before cancer was similarly predictive of advanced and nonadvanced cancers. When AI score was incorporated with breast density measures, the AUC improved for both cancer types. It will be important to evaluate the contribution of the AI scores to the new BCSC model for advanced cancer<sup>22</sup> that includes age, breast density, race/ethnicity, first-degree family history of breast cancer, history of benign breast biopsy, BMI, menopausal status, and screening interval, with the goal of improving identification of those most needing supplemental or alternative screening strategies to biennial mammographic screening. With the already high AUCs for advanced cancer seen in our study (ie, AUC = 0.679) for AI score with only DV, BMI, and age (albeit

**TABLE 3.** Association of Transpara AI Score and Breast Density Measures With Invasive Breast Cancer and With Screen-Detected and Interval Cancers and Advanced and Nonadvanced Cancers

Breast Density Measure	Model <sup>b</sup> With Density Measure Only		Model <sup>b</sup> With Density Measure + Transpara AI Score			Comparison of Models <sup>b</sup> With and Without Transpara AI Score	
	OR Density Measure (95% CI)	AUC (95% CI)	OR Density Measure (95% CI)	OR Transpara AI Score (95% CI)	AUC (95% CI)	Δ AUC (95% CI) P	LRT (P)
All invasive cancer (2,412 cases/ 4,995 controls)							
VPD % <sup>a</sup>	1.47 (1.37 to 1.58)	0.583 (0.569 to 0.597)	1.33 (1.23 to 1.43)	1.18 (1.16 to 1.21)	0.648 (0.634 to 0.661)	Δ 0.064 (0.046 to 0.082) <.001	284.6 (<.001)
DV <sup>a</sup>	1.45 (1.37 to 1.53)	0.587 (0.573 to 0.600)	1.34 (1.26 to 1.42)	1.18 (1.16 to 1.20)	0.648 (0.635 to 0.661)	Δ 0.060 (0.043 to 0.076) <.001	272.0 (<.001)
BI-RADS density <sup>c</sup>							
a	0.62 (0.52 to 0.74)	0.585 (0.572 to 0.599)	0.78 (0.65 to 0.94)	1.19 (1.16 to 1.21)	0.645 (0.632 to 0.660)	Δ 0.059 (0.039 to 0.079) <.001	277.7 (<.001)
b	1.0 (ref)		1.0 (ref)				
c	1.55 (1.37 to 1.75)		1.39 (1.22 to 1.58)				
d	2.24 (1.85 to 2.71)		2.00 (1.64 to 2.43)				
Screen-detected (1,670 cases/ 3,567 controls)							
VPD % <sup>a</sup>	1.35 (1.24 to 1.47)	0.579 (0.563 to 0.595)	1.20 (1.10 to 1.31)	1.20 (1.17 to 1.23)	0.652 (0.636 to 0.668)	Δ 0.072 (0.050 to 0.094) <.001	229.1 (<.001)
DV <sup>a</sup>	1.41 (1.31 to 1.50)	0.585 (0.569 to 0.602)	1.28 (1.19 to 1.38)	1.19 (1.16 to 1.22)	0.650 (0.635 to 0.666)	Δ 0.065 (0.046 to 0.084) <.001	209.3 (<.001)
BI-RADS density							
a	0.60 (0.49 to 0.73)	0.592 (0.576 to 0.608)	0.76 (0.61 to 0.93)	1.19 (1.16 to 1.22)	0.654 (0.638 to 0.669)	Δ 0.061 (0.040 to 0.083) <.001	212.9 (<.001)
b	1.0 (ref)		1.0 (ref)				
c	1.45 (1.25 to 1.68)		1.31 (1.12 to 1.53)				
d	1.88 (1.48 to 2.39)		1.65 (1.29 to 2.12)				

(continued on following page)

**TABLE 3.** Association of Transpara AI Score and Breast Density Measures With Invasive Breast Cancer and With Screen-Detected and Interval Cancers and Advanced and Nonadvanced Cancers (continued)

Breast Density Measure	Model <sup>b</sup> With Density Measure Only		Model <sup>b</sup> With Density Measure + Transpara AI Score			Comparison of Models <sup>b</sup> With and Without Transpara AI Score	
	OR Density Measure (95% CI)	AUC (95% CI)	OR Density Measure (95% CI)	OR Transpara AI Score (95% CI)	AUC (95% CI)	Δ AUC (95% CI) P	LRT (P)
Interval cancer (286 cases/599 controls)							
VPD % <sup>a</sup>	2.36 (1.87 to 2.97)	0.671 (0.633 to 0.709)	2.23 (1.76 to 2.83)	1.17 (1.10 to 1.25)	0.699 (0.663 to 0.736)	Δ 0.027 (-0.015 to 0.070) .20	26.5 (<.001)
DV <sup>a</sup>	1.73 (1.45 to 2.06)	0.646 (0.608 to 0.684)	1.68 (1.40 to 2.01)	1.19 (1.11 to 1.26)	0.694 (0.658 to 0.731)	Δ 0.043 (-0.007 to 0.092) .09	31.4 (<.001)
BI-RADS <sup>c</sup> density							
a	0.47 (0.23 to 0.95)	0.659 (0.621 to 0.697)	0.58 (0.28 to 1.18)	1.18 (1.10 to 1.25)	0.683 (0.645 to 0.720)	Δ 0.032 (-0.019 to 0.084) .21	27.0 (<.001)
b	1.0 (ref)		1.0 (ref)				
c	2.36 (1.59 to 3.51)		2.15 (1.43 to 3.23)				
d	4.62 (2.64 to 8.06)		4.33 (2.43 to 7.70)				
Advanced cancer (289 cases/585 controls)							
VPD % <sup>a</sup>	1.53 (1.26 to 1.86)	0.602 (0.562 to 0.641)	1.34 (1.08 to 1.65)	1.21 (1.14 to 1.29)	0.662 (0.623 to 0.700)	Δ 0.067 (0.018 to 0.116) .007	41.3 (<.001)
DV <sup>a</sup>	1.66 (1.40 to 1.96)	0.624 (0.585 to 0.663)	1.54 (1.29 to 1.83)	1.21 (1.13 to 1.29)	0.679 (0.641 to 0.716)	Δ 0.065 (0.015 to 0.114) .01	39.3 (<.001)
BI-RADS <sup>c</sup> density							
a	0.44 (0.25 to 0.77)	0.603 (0.564 to 0.643)	0.61 (0.34 to 1.09)	1.21 (1.13 to 1.29)	0.646 (0.607 to 0.684)	Δ 0.045 (-0.006 to 0.097) .09	37.2 (<.001)
b	1.0 (ref)		1.0 (ref)				
c	1.62 (1.15 to 2.30)		1.50 (1.04 to 2.14)				
d	1.86 (1.07 to 3.24)		1.31 (0.73 to 2.38)				

(continued on following page)

**TABLE 3.** Association of Transpara AI Score and Breast Density Measures With Invasive Breast Cancer and With Screen-Detected and Interval Cancers and Advanced and Nonadvanced Cancers (continued)

Breast Density Measure	Model <sup>b</sup> With Density Measure Only		Model <sup>b</sup> With Density Measure + Transpara AI Score			Comparison of Models <sup>b</sup> With and Without Transpara AI Score	
	OR Density Measure (95% CI)	AUC (95% CI)	OR Density Measure (95% CI)	OR Transpara AI Score (95% CI)	AUC (95% CI)	Δ AUC (95% CI) P	LRT (P)
Nonadvanced cancer (2,012 cases/4,204 controls)							
VPD % <sup>a</sup>	1.44 (1.33 to 1.56)	0.583 (0.568 to 0.598)	1.31 (1.21 to 1.42)	1.18 (1.15 to 1.21)	0.645 (0.630 to 0.659)	Δ 0.063 (0.043 to 0.082) <.001	229.2 (<.001)
DV <sup>a</sup>	1.40 (1.32 to 1.49)	0.581 (0.566 to 0.596)	1.30 (1.22 to 1.39)	1.18 (1.15 to 1.20)	0.644 (0.630 to 0.659)	Δ 0.060 (0.041 to 0.079) <.001	219.5 (<.001)
BI-RADS <sup>c</sup> density							
a	0.65 (0.53 to 0.78)	0.586 (0.571 to 0.601)	0.81 (0.66 to 0.99)	1.18 (1.16 to 1.21)	0.646 (0.632 to 0.661)	Δ 0.059 (0.038 to 0.080) <.001	227.6 (<.001)
b	1.0 (ref)		1.0 (ref)				
c	1.54 (1.35 to 1.77)		1.39 (1.21 to 1.60)				
d	2.26 (1.84 to 2.78)		2.09 (1.69 to 2.59)				

NOTE. All models adjusted for age and BMI.<sup>a</sup>

Abbreviations: AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System; DV, dense volume; LRT, likelihood ratio test; OR, odds ratio; ref, reference; SD, standard deviation; VPD, volumetric percent density.

<sup>a</sup>ORs represent a change per 1 SD on the log-transformed scale.

<sup>b</sup>ORs for age (per year) ranged from 1.02 (0.99-1.06) to 1.03 (0.99-1.06), and ORs for BMI (per 1 kg/m<sup>2</sup>) ranged from 1.00 (0.99-1.01) to 1.05 (1.04-1.06). BMI was not significant in models with DV.

<sup>c</sup>n = 2,351 cases/4,895 controls for overall BI-RADS analyses (n = 1,626/3,496 for screen-detected; n = 275/584 for interval, and 279/570 for advanced cancer analyses and n = 1,965/4,128 for nonadvanced cancer analyses).

with matching), we expect strong discrimination of advanced cancer with the addition of these other model factors.

Although the AI score significantly increased prediction across all types of cancer, there was less evidence for improvement in discrimination of interval cancers with AI score added to density measures. The smaller gains in AUC likely reflect the strong association of breast density with interval cancers, even when assessed years before the cancers.<sup>15</sup> A prior report<sup>8</sup> showed that inclusion of the AI score with density measures assessed within 2 years of the cancer resulted in improved discrimination of interval cancer but differed from our study in its focus on near versus long-term prediction, its definition of interval cancers over a 2-year versus 1-year period, and use of area-based density<sup>23</sup> versus clinical BI-RADS density or volumetric measures, which may show stronger associations with interval cancer than area-based density measures. Both reports suggest, though, that breast density measures may be complementary to AI detection and risk and need to be considered for prediction of all invasive cancer types.

With the advent of new AI and other imaging risk factors for breast cancer risk, it is important to consider clinically versus statistically significant improvement in discrimination for risk prediction. We define clinically important changes in AUC using prior work on the addition of new risk factors, including benign breast disease (BBD) and the polygenic risk score (PRS), to established risk models for invasive breast cancer. Inclusion of BBD to the BCSC risk model resulted in a change in AUC of 0.65 to 0.66,<sup>24</sup> and inclusion of the breast cancer PRS to the BCSC model resulted in a change in AUC from 0.66 to 0.69.<sup>25</sup> Thus, we propose a change in AUC of 0.01 to 0.03 as clinically significant. Importantly, the inclusion of the Transpara AI score to all models with breast density surpasses this threshold across all invasive cancer types.

Finally, our study cannot disentangle whether the AI score on mammograms 2-5.5 years before the cancer is reflective of suspicious findings underlying breast cancers missed at review or identifying suspicious changes that may develop into a detectable breast cancer in the future. Our models that incorporate both BI-RADS assessment and the AI score show

that the AI score is robust and contributes additional risk information not contained in the BI-RADS assessment. Importantly, some of the high AI scores, reflective of suspicious changes, are subsequently associated with a diagnosis of advanced cancer or aggressive tumors that are not detectable until rapid growth into a detectable mass. It is an advantage of AI algorithms to identify suspicious areas that could develop into advanced cancer since it provides the opportunity to apply screening strategies among those at high risk of these cancers to potentially increase detection of these tumors at an earlier stage.

We acknowledge several limitations in our study including focus on two-dimensional mammography instead of digital breast tomosynthesis or three-dimensional mammography, the focus on one AI algorithm, and the limited power to evaluate groups by race or ethnicity. In particular, our statistical power for comparisons within advanced cancer and interval cancer groups might have limited our ability to detect smaller contributions to discriminatory accuracy of imaging-based AI score to density models. Although racial and ethnic differences have not been seen to date for Mirai,<sup>1</sup> there have been differences seen in volumetric density by race,<sup>26</sup> and it will be important to consider future analyses within multiple races and ethnicities. In addition, the retrospective and case/control nature of the study design limits absolute risk estimation and enriches for cancer, not allowing for actual estimates for the impact of AI when in wider use; prospective cohort studies will allow the estimation of absolute as opposed to relative risk and evaluate the impact of the AI score on relevant clinical thresholds.

In conclusion, the Transpara AI score improves long-term risk prediction when combined with clinical risk factors including breast density for overall invasive cancers, screen-detected, advanced, and nonadvanced cancers. For interval cancers, density measures remain of greatest importance for discrimination, even years before the cancer. Incorporating both density and AI scores into long-term risk models will be important for accurate prediction of invasive cancer outcomes, in particular, advanced cancer.

## AFFILIATIONS

<sup>1</sup>Division of Epidemiology, Department Quantitative Sciences, Mayo Clinic, Rochester, MN

<sup>2</sup>Division of Clinical Trials and Biostatistics, Department Quantitative Sciences, Mayo Clinic, Rochester, MN

<sup>3</sup>Division of Breast Imaging, Department of Radiology, Mayo Clinic, Rochester, MN

<sup>4</sup>Division of Computational Biology, Department Quantitative Sciences, Mayo Clinic, Rochester, MN

<sup>5</sup>University of California, San Francisco, CA

<sup>6</sup>Department Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN

## CORRESPONDING AUTHOR

Celine M. Vachon, PhD, Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic, 200 First St SW, Biobusiness BB 5-8, Rochester, MN 55905; e-mail: vachon.celine@mayo.edu.

## DISCLAIMER

The funding source had no role in the conduct of the research and/or preparation of the article, or in the study design; collection, analysis, or interpretation of the data, or in the writing of the report; or in the decision to submit the article for publication.

**SUPPORT**

This research was made possible by funding from the National Institutes of Health, grant numbers P01CA154292, R01 CA177150, R01 CA207084, and R01 CA275074.

**AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

Disclosures provided by the authors are available with this article at DOI <https://doi.org/10.1200/JCO.22.01153>.

**AUTHOR CONTRIBUTIONS**

**Conception and design:** Celine M. Vachon, Aaron D. Norman, Stacey J. Winham, Karla Kerlikowske

**Financial support:** Celine M. Vachon, Karla Kerlikowske

**Administrative support:** Aaron D. Norman, Karla Kerlikowske

**Provision of study materials or patients:** Celine M. Vachon, Karla Kerlikowske

**Collection and assembly of data:** Celine M. Vachon, Christopher G. Scott, Aaron D. Norman, Matthew R. Jensen, Karla Kerlikowske

**Data analysis and interpretation:** Celine M. Vachon, Christopher G. Scott, Sadia A. Khanani, Matthew R. Jensen, Carrie B. Hruska, Kathleen R. Brandt, Stacey J. Winham, Karla Kerlikowske

**Manuscript writing:** All authors

**Final approval of manuscript:** All authors

**Accountable for all aspects of the work:** All authors

**ACKNOWLEDGMENT**

We thank Professor Nico Karssemeijer (ScreenPoint Medical) for providing the Transpara AI software and all the women who provide research authorization to allow their medical records to be used for research.

**REFERENCES**

1. Yala A, Mikhael PG, Strand F, et al: Multi-institutional validation of a mammography-based breast cancer risk model. *J Clin Oncol* 40:1732-1740, 2022
2. Eriksson M, Czene K, Strand F, et al: Identification of women at high risk of breast cancer who need supplemental screening. *Radiology* 297:327-333, 2020
3. Rodriguez-Ruiz A, Krupinski E, Mordang JJ, et al: Detection of breast cancer with mammography: Effect of an artificial intelligence support system. *Radiology* 290:305-314, 2019
4. Rodriguez-Ruiz A, Lang K, Gubern-Merida A, et al: Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. *J Natl Cancer Inst* 111:916-922, 2019
5. Kerlikowske K, Zhu W, Tosteson AN, et al: Identifying women with dense breasts at high risk for interval cancer: A cohort study. *Ann Intern Med* 162:673-681, 2015
6. Mandelson MT, Oestreicher N, Porter PL, et al: Breast density as a predictor of mammographic detection: Comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 92:1081-1087, 2000
7. Boyd NF, Guo H, Martin LJ, et al: Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 356:227-236, 2007
8. Wanders AJT, Mees W, Bun PAM, et al: Interval cancer detection using a neural network and breast density in women with negative screening mammograms. *Radiology* 303:269-275, 2022
9. Kerlikowske K, Bissell MCS, Sprague BL, et al: Advanced breast cancer definitions by staging system examined in the Breast Cancer Surveillance Consortium. *J Natl Cancer Inst* 113:909-916, 2021
10. Kerlikowske K, Chen S, Golmakan MK, et al: Cumulative advanced breast cancer risk prediction model developed in a screening mammography population. *J Natl Cancer Inst* 114:676-685, 2022
11. Breast Cancer Surveillance Consortium: <http://www.bsc-research.org/>
12. Pinto MC, Rodriguez-Ruiz A, Pedersen K, et al: Impact of artificial intelligence decision support using deep learning on breast cancer screening interpretation with single-view wide-angle digital breast tomosynthesis. *Radiology* 300:529-536, 2021
13. Larsen M, Aglen CF, Lee CI, et al: Artificial intelligence evaluation of 122 969 mammography examinations from a population-based screening program. *Radiology* 303:502-511, 2022
14. Brandt KR, Scott CG, Ma L, et al: Comparison of clinical and automated breast density measurements: Implications for risk prediction and supplemental screening. *Radiology* 279:710-719, 2016
15. Kerlikowske K, Scott CG, Mahmoudzadeh AP, et al: Automated and clinical breast imaging reporting and data system density measures predict risk for screen-detected and interval cancers: A case-control study. *Ann Intern Med* 168:757-765, 2018
16. Brentnall AR, Cuzick J, Field J, et al: A concordance index for matched case-control studies with applications in cancer risk. *Stat Med* 34:396-405, 2015
17. Pepe MS, Kerr KF, Longton G, et al: Testing for improvement in prediction model performance. *Stat Med* 32:1467-1482, 2013
18. Yala A, Mikhael PG, Strand F, et al: Toward robust mammography-based models for breast cancer risk. *Sci Transl Med* 13:eaba4373, 2021
19. Dembrower K, Liu Y, Azizpour H, et al: Comparison of a deep learning risk score and standard mammographic density score for breast cancer risk prediction. *Radiology* 294:265-272, 2020
20. Dembrower K, Wahlin E, Liu Y, et al: Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: A retrospective simulation study. *Lancet Digital Health* 2:e468-e474, 2020
21. Kopans DB: Design, implementation, and pitfalls of TMIST. *Clin Imaging* 78:304-307, 2021
22. Kerlikowske K, Carney PA, Geller B, et al: Performance of screening mammography among women with and without a first-degree relative with breast cancer. *Ann Intern Med* 133:855-863, 2000
23. Gastouniotti A, Kasi CD, Scott CG, et al: Evaluation of LIBRA software for fully automated mammographic density assessment in breast cancer risk prediction. *Radiology* 296:24-31, 2020
24. Tice JA, Miglioretti DL, Li CS, et al: Breast density and benign breast disease: Risk assessment to identify women at high risk of breast cancer. *J Clin Oncol* 33:3137-3143, 2015
25. Vachon CM, Pankratz VS, Scott CG, et al: The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst* 107:dju397, 2015
26. McCarthy AM, Keller BM, Pantalone LM, et al: Racial differences in quantitative measures of area and volumetric breast density. *J Natl Cancer Inst* 108:djw104, 2016



**AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

**Impact of Artificial Intelligence System and Volumetric Density on Risk Prediction of Interval, Screen-Detected, and Advanced Breast Cancer**

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to [www.asco.org/rwc](http://www.asco.org/rwc) or [ascopubs.org/jco/authors/author-center](http://ascopubs.org/jco/authors/author-center).

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](#)).

**Celine M. Vachon**

**Research Funding:** GRAIL (Inst)

**Patents, Royalties, Other Intellectual Property:** Breast Density software

**Sadia A. Khanani**

**Research Funding:** Siemens Healthineers

**Patents, Royalties, Other Intellectual Property:** Mayo/SK and Imago Systems have a research and development agreement (Inst)

**Carrie B. Hruska**

**Research Funding:** Siemens (Inst)

**Patents, Royalties, Other Intellectual Property:** Receive royalties for MBI technologies licensed to CMR Naviscan, Licensing agreement with CMR Naviscan (Inst)

No other potential conflicts of interest were reported.