Title
Evaluation of Breast Imaging Biomarkers for Breast Cancer Future Risk and Treatment Response

Author
Arasu, Vignesh Amal

Publication Date
2021

Peer reviewed|Thesis/dissertation

Evaluation of Breast Imaging Biomarkers for Future Cancer Risk and Treatment Response

by
Vignesh A. Arasu

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Epidemiology and Translational Science

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Nola M. Hylton
D750F046316D4D6...

Nola M. Hylton
Chair

DocuSigned by:
Maria Glymour
Doc[illegible]41C...

Maria Glymour

DocuSigned by:
Laura Esserman
Doc[illegible]48B...

Laura Esserman

DocuSigned by:
John Kornak
Doc[illegible]4CD...

John Kornak

DocuSigned by:
Diana L Miglioretti
85CBFABE29EE456...

Diana L Miglioretti

Committee Members

Dedication and Acknowledgments

This work is dedicated to my family – my wife, son, parents, and sister – for supporting me for the years I have invested in this body of work. I am grateful for their belief in me, as it was by no means a fun process for them: time and cost of adding a separate degree above and beyond my normal day job, irritation with my many moments of preoccupation with research, and anguish when I had to work late nights and early mornings. This PhD and dissertation work also occurred alongside many life events, including completing residency and fellowship, finding a job, getting married, purchasing our first home, having a baby, and facing the passing of loved ones. But their underlying patience and confidence that I was doing the right thing helped me overcome my own doubts through the process.

To mentors and collaborators, old and new, who have given me this incredible gift of knowledge and better understanding of how to think. Specifically, my dissertation committee, Drs. Nola Hylton, Maria Glymour, Laura Esserman, John Kornak, and Diana Miglioretti. My KP research collaborators Drs. Laurel Habel and Catherine Lee. My prior UCSF radiology mentors Drs. Bonnie Joe and Edward Sickles. More removed, but nonetheless very impactful for setting me up to succeed in a research career, my undergraduate mentors Drs. Nicole Weekes and Richard Lewis, my "gap year" research mentor Dr. Calvin Hobel, and summer mentor Dr. Margaret Bauman. And finally to my grade school teachers, Mr. Adams, Mr. Wilmot, and Mr. Nelson.

To others I worked alongside with who helped stimulate my development and this work, including my PhD classmates Michelle Roh, Luis Rodriguez, and Rae Wannier. At UCSF Hylton

Lab, Wen Li, David Newitt, Ella Jones, and Lisa Wilmes. At UCSF/BCSC, Karla Kerlikowskee and Jeff Tice. At KP, Ninah Achacoso, the Physician Researcher Program, and Caitlin Lydon. My prior mentees who helped me learn through the process of teaching, Paul Kim and Cody McHargue. This work depended on many people I will never know but nonetheless thankful for their involvement, including the many women participating in these studies, radiologists, and imaging facilities.

Chapter 1 was financially supported by National Institutes of Health U01 CA225427 and National Institutes of Health R01 CA132870, Quantum Leap Healthcare Collaborative, Foundation for the National Institutes of Health, National Cancer Institute (Grant 28XS197 P-0518835), Safeway, an Albertsons Company, William K. Bowes, Jr. Foundation, Breast Cancer Research Foundation, UCSF, GMU, the Biomarkers Consortium, Salesforce, OpenClinica, Formedix, Natera, Hologic Inc., TGen, Illumina, CCS Associates, Berry Consultants, Breast Cancer Research—Atwater Trust, Stand up to Cancer, California Breast Cancer Research Program, and Give Breast Cancer the Boot, IQVIA, Genentech, Amgen, Pfizer, Merck, Seattle Genetics, Daiichi Sankyo, AstraZeneca, Dynavax Technologies, Puma Biotechnology, AbbVie, Madrigal Pharmaceuticals (formerly Synta Pharmaceuticals), Plexxikon, Regeneron, and Agendia. Chapter 2 was supported by National Cancer Institute–funded Breast Cancer Surveillance Consortium program project (Grant No. P01CA154292), National Cancer Institute Grant No. U54CA163303, Patient-Centered Outcomes Research Institute Grant No. PCS-1504-30370, and Agency for Healthcare Research and Quality Grant No. R01 HS018366-01A1, and several state public health departments and cancer registries throughout the United States. Chapter 3 was supported by the Kaiser Permanente Delivery and Applied Science Program.

Contributions

A version of Chapter 1 in this dissertation was published in The Journal of Breast Imaging in August 2020, and a version of Chapter 2 was published in The Journal of Clinical Oncology on April 2019. The Dissertation Committee Members supervised the research that forms the basis of these dissertation chapters. The published material is substantially the product of Vignesh Arasu's period of study at the University of California, San Francisco and was primary conducted and written by him. The work he completed for this published manuscript is comparable to a standard dissertation chapter.

Epigraph

“Go see the Muni train.”

-Kiran Shah-Arasu

“Matha, Pitha, Guru, Devam”

- Sanskrit Proverb

Evaluation of Breast Imaging Biomarkers for Breast Cancer Future Risk and Treatment Response

Vignesh Arasu

Abstract

Imaging biomarkers are representations of an *in vivo* biological state and phenotype. The incorporation of breast density in breast cancer risk models, as well as state-mandated reporting of mammographic breast density to women, underscores the central role of imaging biomarkers in risk assessment. In this dissertation, I evaluate breast imaging biomarkers from breast MRI and mammography in their role of future risk prediction and treatment response. The chapters, ordered chronologically, show the evolution of my research interests from quantitative imaging science within a well-controlled experimental trial (Chapter 1), to a population-based evaluation of qualitative clinically derived imaging assessments in an observational cohort (Chapter 2), to finally combining quantitative imaging science for comparative evaluations through a population-based pragmatic assessment in a large managed health system (Chapter 3).

Chapters 1 and 2 focus on background parenchymal enhancement (BPE), which describes the natural phenomenon observed on breast MRI in which normal breast tissue demonstrates signal enhancement related to uptake of intravenous contrast. Biologically, BPE is believed to represent tissue "activated" by endogenous hormones (primarily estrogen) and is dynamic in appearance over time and distribution within a woman's breast tissue. Chapter 1 focuses on manually defined quantitative imaging biomarkers in the experimental I-SPY 2 trial, an on-going multicenter prospective randomized clinical trial framework used to monitor treatment response and assess novel investigational neoadjuvant chemotherapy (NAC) agents for breast cancer.

Women with advanced HER2- breast cancer have limited treatment options. Breast MRI functional tumor volume (FTV) is used to predict pathologic complete response (pCR) to improve treatment efficacy. In addition to FTV, background parenchymal enhancement (BPE) may predict response and was explored for HER2- patients in the ISPY-2 TRIAL. We found that among women with HER2- cancer, BPE alone demonstrated association with pCR in women with HR+HER2- breast cancer, with similar diagnostic performance to FTV. BPE predictors remained significant in multivariate FTV models, but without added discrimination for pCR prediction. This may be due to small sample size limiting ability to create subtype specific multivariate models.

Chapter 2 extends BPE evaluation through comparative associations of qualitative BPE and mammographic breast density for future risk in a population-based assessment using the Breast Cancer Surveillance Consortium (BCSC), involving 46 radiology facilities that participate in one of six regional BCSC registries. Higher levels of BPE were found to be associated with future invasive breast cancer risk independent of breast density. The combination of both high BPE and high breast density was associated with higher risk than either factor alone. BPE also demonstrates subtype specific associations with less aggressive disease, although the association with aggressive disease was noted at moderate and marked levels.

Finally, Chapter 3 examines whether using computer vision artificial intelligence (AI)–based computer vision algorithms, most of which are trained to extract features from mammograms to detect visible breast cancer, can also predict future risk using a population-based case cohort from the Kaiser Permanente Northern California managed health system. We found that all AI mammography algorithms evaluated had clinically and statistically significantly higher discrimination than the BCSC clinical risk model for interval cancer and 5-year future

cancer risk, indicating their usefulness. The combination of BCSC and AI further improves risk prediction above AI alone, and decreases the gap in future risk performance between AI algorithms. Training AI algorithms to predict longer-term outcomes may yield further improvements, but the potential impact on clinical decisions requires further study.

Table of Contents

CHAPTER 1: Predictive value of breast MRI background parenchymal enhancement for neoadjuvant treatment response among HER2- patients 1

ABSTRACT 1

INTRODUCTION 3

METHODS 5

RESULTS 9

DISCUSSION 11

REFERENCES 14

CHAPTER 2: Population-Based Assessment of the Association between MRI Background Parenchymal Enhancement and Future Primary Breast Cancer Risk 25

ABSTRACT 25

INTRODUCTION 27

METHODS 28

RESULTS 32

DISCUSSION 37

REFERENCES 41

CHAPTER 3: Comparison of Mammography Artificial Intelligence Algorithms for 5-Year Breast Cancer Risk Prediction 60

ABSTRACT 60

INTRODUCTION 62

METHODS 63

RESULTS 68

DISCUSSION 70

REFERENCES 74

List of Figures

Figure 1.1: I-SPY2 TRIAL schema. 21

Figure 1.2: Process of quantitative background parenchymal enhancement (BPE) calculation .22

Figure 1.3: Plots of median values of background parenchymal enhancement (BPE) and functional tumor volume (FTV) through phases of treatment 23

Figure 1.4: Spectral maximum intensity projection breast MRI of an individual woman's background parenchymal enhancement (BPE).....24

Figure 2.1: Distribution of MRI background parenchymal enhancement (BPE) and mammographic breast density 47

Figure 3.1: Case cohort selection.....90

Figure 3.2: Cumulative risk of breast cancer by AI score decile at 5-years 91

List of Tables

Table 1.1: Participant Characteristics 18

Table 1.2: Univariate analyses of BPE variables for predicting pCR, stratified by HR subtype..19

Table 1.3: Comparison of pathologic complete response (pCR) prediction models based on functional tumor volume (FTV) predictors only or adding background parenchymal enhancement (BPE) predictors.....20

Table 2.1: Women's characteristics at time of first MRI background parenchymal enhancement (BPE) measurement by cancer status and BPE level.....48

Table 2.2: Associations between baseline MRI background parenchymal enhancement (BPE) and breast cancer risk.....51

Table 2.3: Associations among mammographic breast density, BCSC 5-year risk, and MRI background parenchymal enhancement (BPE) on breast cancer risk 52

Table 2.4: Secondary and sensitivity analyses of mild, moderate, or marked versus minimal MRI background parenchymal enhancement (BPE) on breast cancer risk. 53

Table 2.5a: Cancer subgroup analysis by baseline BPE and/or breast density: Invasive cancer and DCIS 54

Table 2.5b: Cancer subgroup analysis by baseline BPE and/or breast density: by HR and HER2 status 55

Table 2.5c: Cancer subgroup analysis by baseline BPE and/or breast density: By stage.....56

Table 2.6a: Cancer subgroup analysis by multiple BPE and/or multiple breast density measurements: Invasive cancer and DCIS 57

Table 2.6b: Cancer subgroup analysis by multiple BPE and/or multiple breast density measurements: by HR and HER2 status 58

Table 2.6c: Cancer subgroup analysis by multiple BPE and/or multiple breast density measurements 59

Table 3.1: Missing data by model 79

Table 3.2: Complete cases across all models only 80

Table 3.3: Baseline patient characteristics 81

Table 3.4: Comparative AUC(t) performance of AI models and BCSC clinical risk model using 2016 screening mammogram without evidence of cancer on final imaging assessment for invasive cancer and DCIS 82

Table 3.5: Comparative time-varying AUC(t) performance of combined AI models and BCSC clinical risk model using 2016 screening mammogram without evidence of cancer on final imaging assessment 83

Table 3.6: Comparative AUC(t) performance of AI models and BCSC clinical risk model for invasive cancer only using 2016 screening mammogram without evidence of cancer on final imaging assessment 84

Table 3.7: Screening mammograms acquired on Hologic equipment, complete cases only 85

Table 3.8: Screening mammograms acquired on General Electric equipment, complete cases only 86

Table 3.9: Screening BI-RADS 1 or 2 only 87

Table 3.10: Screening BI-RADS 0, and diagnostic or post-procedural BI-RADS 1 or 2 only ... 88

Table 3.11: Absolute risk calibration by Mirai AI and BCSC clinical risk models at 5-year cumulative incidence thresholds (using test set only) 89

List of Abbreviations

AC	Anthracycline-cyclophosphamide
AI	Artificial intelligence
AJCC	American Joint Committee on Cancer
AUC	Area under the curve
BCRAT	Breast Cancer Risk Assessment Tool
BCSC	Breast Cancer Surveillance Consortium
BPE	Background parenchymal enhancement
CI	Confidence interval
CNN	Convolutional neural networks
CV	Cross-validation
DCE	Dynamic contrast-enhanced
DCIS	Ductal carcinoma in situ
FTV	Functional tumor volume
GMIC	Globally Aware Multiple Instance Classifier
HR	Hazard ratio
IQR	Inter-quartile range
IR	Incidence rates
IRR	Incidence rate ratio
KPNC	Kaiser Permanente Northern California
MRI	Magnetic resonance imaging
NAC	Neoadjuvant chemotherapy
NYU	New York University

OR	Odds ratio
SEER	Surveillance, Epidemiology, and End Results
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology

CHAPTER 1: Predictive value of breast MRI background parenchymal enhancement for neoadjuvant treatment response among HER2- patients

ABSTRACT

Objective

Women with advanced HER2- breast cancer have limited treatment options. Breast MRI functional tumor volume (FTV) is used to predict pathologic complete response (pCR) to improve treatment efficacy. In addition to FTV, background parenchymal enhancement (BPE) may predict response and was explored for HER2- patients in the ISPY-2 TRIAL.

Methods

Women with HER2- stage II or III breast cancer underwent prospective serial breast MRIs during four neoadjuvant chemotherapy timepoints. BPE was quantitatively calculated using whole breast manual segmentation. Logistic regression models were systematically explored using pre-specified and optimized predictor selection based on BPE or combined with FTV.

Results

A total of 352 MRI examinations in 88 patients (29 with pCR, 59 non-pCR) were evaluated. Women with HR+HER2- cancers who achieved pCR demonstrated a significantly greater decrease in BPE from baseline to pre-surgery compared to non-pCR patients (OR 0.64, 95%CI 0.39-0.92, *p*-value = 0.04). The associated BPE area under the curve (AUC) was 0.77 (95%CI 0.56-0.98), comparable to the range of FTV AUC estimates. Among multi-predictor models, the highest cross-validated AUC of 0.81 (95%CI 0.73-0.90) was achieved with combined FTV+HR

predictors, while adding BPE to FTV+HR models had an estimated AUC of 0.82 (95%CI 0.74-0.92).

Conclusion

Among women with HER2- cancer, BPE alone demonstrated association with pCR in women with HR+HER2- breast cancer, with similar diagnostic performance to FTV. BPE predictors remained significant in multivariate FTV models, but without added discrimination for pCR prediction. This may be due to small sample size limiting ability to create subtype specific multivariate models.

INTRODUCTION

Women with advanced breast cancer (stage 2 and 3) have significant morbidity and mortality, with a 5-year disease specific survival as low as 33%.[1] The neoadjuvant period provides the opportunity to noninvasively monitor tumor response to therapy with breast MRI, and redirect therapy for women who are not responding in hopes of improving their prognosis. Furthermore, the surrogate outcome pathologic complete response (pCR) has a high association with survival, accelerating the prediction of a woman's outcome to months rather than years.[2] Women with advanced HR+HER2- and HR-HER2- disease in particular have relatively lower rates of pCR as compared to women HER2+ due to limited treatment options.[3] Improving prediction of pCR in women with HER2- disease during the neoadjuvant period would provide opportunities to improve treatment selection and potentially increase the pCR rate.

The I-SPY 2 TRIAL (Investigation of Serial Studies to Predict Your Therapeutic Response through Imaging and Molecular Analysis 2, clinicalTrials.gov number NCT01042379) is an on-going multicenter prospective randomized clinical trial framework used to monitor treatment response and assess novel investigational neoadjuvant chemotherapy (NAC) agents for breast cancer. The study uses quantitative measurement of magnetic resonance imaging (MRI)-derived tumor volume (defined as functional tumor volume or FTV) to predict response. The prior ISPY-1 trial demonstrated a significant association with both prediction of pCR[4] and recurrence-free survival[5] outcomes, with area under the curve (AUC) estimates for FTV regression models ranging 0.70-0.84 and 0.52-0.72 for each outcome type respectively.

Background parenchymal enhancement (BPE) describes the natural phenomenon observed on breast MRI in which normal breast tissue demonstrates signal enhancement related to uptake of intravenous contrast. Biologically, BPE is believed to represent tissue "activated" by endogenous hormones (primarily estrogen), and is dynamic in appearance over time and distribution within a woman's breast tissue. This is demonstrated by histopathologic studies that have found BPE to be correlated with increased microvascular density[6] and proliferative breast tissue[7]. Additionally, single-center studies have found strong associations between BPE and subsequent primary breast cancer, with odds ratios of 2-18.[8-10] More recent studies have also demonstrated that BPE is a surrogate outcome of treatment response to chemotherapy and chemoprevention agents[11-14]. BPE signal intensity decreases with treatment, and the magnitude of this decrease is associated with the degree of tumor response. The biological basis of these associations are unclear, but it has been speculated that BPE characterizes "activated" breast stroma that is more susceptible to malignant transformation but also to potential treatment responsiveness.[9,15]

Tumor volume is a validated predictor of NAC response, but few MRI studies have evaluated the adjunctive contribution of BPE to a tumor volume model. Most studies evaluate the association of BPE alone with treatment response, but the more relevant clinical question is if BPE provides additive improvement to the more established tumor volume model. Moreover, prior studies on BPE in tumor response are based on retrospective observational studies from single institutions or rely on a qualitative definition of BPE, which is prone to issues with inter-rater reliability and measurement error[12,16].

The current study has several strengths that overcome limitations in prior studies: we analyze data from a prospective study primarily designed to evaluate MR imaging biomarkers; our patient cohort was evaluated for a clearly defined pathological endpoint for neoadjuvant response; we had consistent MRI protocol with high quality control of acquisition. We evaluated the primary effect of BPE as well as the additive effect of BPE to a FTV tumor volume model in improving the prediction of pCR of women with HER2- advanced breast cancer enrolled in the ISPY 2 trial.

METHODS

Patient Population

In this Health Insurance Portability and Accountability Act-compliant, Institutional Review Board-approved study, women 18 years of age and older diagnosed with stage II or III breast cancer and with tumor size measured ≥ 2.5 cm were eligible to enroll in the I-SPY 2 TRIAL[17]. Biomarker assessments based on hormone (estrogen and progesterone) receptors (HR+/-) and a 70-gene assay (MammaPrint, Agendia, Amsterdam, The Netherlands) were performed at baseline and used for treatment randomization[17]. Patients who had tumors that were designated as hormone-receptor positive and low risk according to the 70-gene assay were excluded. All patients provided written informed consent to participate in the study. A second consent was obtained if the patient was randomized to an experimental treatment. Enrollment occurred between 2010-2012.

Schema

Figure 1.1 shows the schema of the I-SPY 2 TRIAL. All breast cancers in these drug arms were HER2 negative by nature of the drug mechanism of action. Participants received a weekly dose of paclitaxel alone (control) or in combination with experimental NAC agents for 12 weekly cycles, followed by four (every 2-3 weeks) cycles of anthracycline-cyclophosphamide (AC) prior to surgery. MRI was performed before the initiation of NAC or "baseline" (T0), after three weeks of therapy or "early treatment" (T1), after twelve weeks of therapy at which patient is transitioned from taxane-based regimen to AC-based regimen or "inter-regimen" (T2), and after neoadjuvant therapy completion and prior to surgery or "pre-surgery" (T3).

Pathologic Assessment of Response

Pathologic complete response — defined as the absence of residual invasive cancer in the breast or lymph nodes at the time of surgery — is the primary end point of the I-SPY 2 TRIAL. All patients were classified as achieving pCR or not achieving pCR (non-pCR) at the time of definitive surgery by a trained pathologist. Patients that withdrew from the trial in mid-study were counted as non-pCR.

MRI Acquisition

MR imaging was performed using 1.5T or 3T scanners with a dedicated breast radiofrequency coil, across a variety of vendor platforms and institutions. All MRI exams within a single patient were performed using the same magnet configuration (manufacturer; field strength; breast coil model). Bilateral dynamic contrast-enhanced (DCE) MRI images were acquired in the axial orientation with the following parameters: TR = 4–10 ms, minimum TE, flip angle = 10–20

degrees, field of view (FOV) = 260–360 mm to achieve full bilateral coverage, acquisition matrix = 384–512 within-plane resolution ≤ 1.4 mm, slice thickness ≤ 2.5 mm, and slice gap = 0 mm. Gadolinium contrast agent was administrated intravenously at a dose of 0.1 mmol/kg body weight, and at a rate of 2 mL/second, followed by a 20 mL saline flush. The same contrast agent brand was used for all MRI exams for the same patient. Pre-contrast and multiple post-contrast images were acquired using identical sequence parameters. There was no delay between contrast injection and data acquisition. Post-contrast imaging continued for at least 8 minutes following contrast agent injection.

Quantitative Image Analysis

FTV was calculated from each DCE-MRI examination using a previously described semi-automated segmentation method (Figure 1.2)[18]. BPE was assessed following manual whole breast segmentation of the contralateral unaffected breast so that measurement would not be confounded by adjacent disease. Subsequently, enhancement was determined on a per-voxel basis using co-registered DCE sequences at two time points: pre-contrast (time 0) and the first post-contrast acquisition between 2 minutes 15 seconds and 2 minutes 30 seconds post-contrast at (time 1), with S_0 and S_1 representing the corresponding signal intensities at those times. BPE was calculated as an average of early enhancement measured for all voxel of segmented fibroglandular tissue, where early enhancement is defined as $(S_1 - S_0)/S_0$.

For FTV measurements, the segmentation method calculated the volume of all tumor voxels that exceeded an early enhancement threshold of 70%. Participating sites in I-SPY 2 TRIAL could slightly adjust the early enhancement threshold to qualitatively reflect the extent of tumor, and to

account for unexpected variability in MRI systems and imaging parameters. However, the FTV analysis had to be reviewed and approved by the designated breast radiologist at each site, and all FTVs in I-SPY 2 TRIAL had to be visually approved by the Imaging Core Lab at the University of California San Francisco.

Statistical Analysis

Univariate analyses were performed with logistic regression, using predictors of absolute values of BPE and FTV at each treatment time point (e.g., absolute value of BPE at inter-regimen/T2 is notated as "BPE_2") or relative change from baseline, and the treatment response outcome pCR. Relative change was calculated as change from baseline divided by baseline value. For example, relative change of BPE from baseline to early treatment (or T1) was calculated as (BPE_1 – BPE_0)/BPE_0 and notated as %ΔBPE0_1. All possible FTV or BPE predictors were evaluated as individual univariate predictors of tumor response in models stratified by HR status. We additionally estimated models including the following sets of multiple predictors: Model 1) baseline FTV and relative FTV change for each treatment time point; Model 2) the same FTV model with the corresponding baseline and BPE change variable. A final model was derived which optimized AUC by exhaustively searching all possible linear combination of FTV predictors and HR, without or with all possible BPE predictors ("Model 3" and "Model 4", respectively). For all models, an odds ratio (OR) are used to describe strength of association with pCR. For the relative change measures ORs are reported for 10% relative change to aid interpretability. The interpretation of the OR, for example an OR of 0.9 for a relative change variable, is that for each 10% decrease in ΔFTV or ΔBPE, there is a 10% decrease in the odds of non-pCR or corresponding 10% increase in the odds of pCR. Diagnostic performance was

assessed using area under the curve (AUC) for all models. To avoid overfitting, 10 times repeated 5-fold cross-validation AUC (cvAUC) was used for multiple predictor models. All statistical analysis was performed using the R statistical programming environment, version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria). A nominal *p*-value of < 0.05 was considered to be statistically significant.

RESULTS

Of the 110 women who had enrolled and received at least one MRI examination in the initial drug arms, a total of 88 women (29 with pCR, 59 with non-pCR) with 352 MRI examinations were included. A total of 22 women were excluded for the following reasons: unable to calculate BPE due to image quality issues (13 women), missing one or more MRI visits (8 women), and missing demographic information (1 woman).

Patient characteristics

Table 1.1 describes the baseline characteristics of women included in this study. Women with pCR as compared to women with non-pCR were slightly younger and more often Asian or Black/African-American and pre-menopausal. Women with pCR were more commonly HR+HER2- than women with non-PCR.

Univariate analysis of BPE

Figure 1.3 displays the average absolute values of BPE and FTV over time as treatment progressed. Women who achieved pCR tended to have higher absolute BPE values at baseline, which decreased more at later treatment time points than non-pCR patients (Figure 1.4). In

contrast, women who achieved pCR tended to have lower absolute FTV values at baseline, which remained lower for all time points than non-pCR patients.

Table 1.2 summarizes our findings of the univariate regression analyses of all 88 women included in this study stratified by HR status. Greater decreases in BPE from baseline to inter-regimen treatment predicted a higher odds of pCR (%ΔBPE0_2; OR = 0.88 per 10% change in %BPE0_2, 95% CI = 0.75 to 1.00, *p*-value = 0.08) or from baseline to pre-surgery (%ΔBPE0_3; OR = 0.87 per 10% change in predictor, 95% CI = 0.74 to 1.00, *p*-value = 0.07) than non-pCR, although the *p*-value and AUC did not reach statistical significance at the nominal $\alpha = 0.05$ level for either predictor. Among the 43 women with HR+ breast cancer, the change in BPE from baseline to pre-surgery was statistically significant (%ΔBPE0_3; OR = 0.64 per 10% change in predictor, 95% CI = 0.39 to 0.92, *p*-value = 0.04), with a corresponding AUC of 0.77 (95% CI 0.56 to 0.98). In comparison, FTV univariate AUCs ranged from 0.57-0.80 in this population depending on the FTV predictor used.

Multiple predictor analysis of BPE and FTV

Table 1.3 describes the results of the multiple predictor analyses, which were used to assess the additive effect of BPE to FTV-only multiple predictor models. FTV-only multiple predictor analyses demonstrated statistically significant associations across subtypes in change parameters only, with cvAUC remaining significant and estimates ranged from 0.61-0.72. Model 2 added BPE to Model 1, which did not lead to improved overall performance based on cvAUC.

Model 3 included all possible linear combination of FTV and HR predictors, which was then optimized by selecting the highest cvAUC value, achieving a cvAUC of 0.81 (95% CI 0.73 to 0.90) with %ΔFTV0_2 and HR status. Model 4 was based on any possible combination of BPE predictors with FTV and HR predictors, which did not substantially change cvAUC of 0.82 (95% CI 0.74 to 0.92) but retained multiple BPE predictors with significant associations with pCR.

DISCUSSION

In this study, we demonstrated that quantitative whole breast BPE alone was predictive of pCR using change from baseline to later treatment time points in women with HR+HER2- breast cancers who were undergoing taxane and anthracycline-based NAC regimen. Moreover, the diagnostic accuracy as measured by AUC was comparable to the predictive performance of the tumor volume measurement FTV. BPE predictors remaining significantly associated with pCR when added to multivariate FTV models, however there was no substantial improvement in discrimination.

We observed that BPE responds to neoadjuvant therapy as demonstrated by declining values as treatment progressed. Moreover, BPE had a similar diagnostic accuracy for women with HR+HER2- breast cancers as compared to FTV under univariate analysis. This is impressive in so far as BPE is measured in the contralateral unaffected breast of presumably normal fibroglandular tissue, whereas FTV is a direct measurement of the primary disease. This suggests that the reaction of normal tissue to neoadjuvant therapy as measured by BPE may represent a biomarker of treatment responsive phenotype, with higher sensitivity to HR+ tumors. This is consistent with the theory of BPE being primarily modulated by estrogen, given higher values in

pre-menopausal women and consistent decreases with hormone therapy[11]. Moreover, this is consistent with multiple prior studies, which demonstrated changes in BPE in response to chemotherapy for prediction of pCR[19]. However, the observed nature of subtype specific effect has been mixed in prior studies, with some studies showing an effect of BPE only in HR+ subtypes[14,20], and some studies showing an effect in HR- subtypes[21,22].

The additive value of BPE remains uncertain based on our multiple predictor results. While retained BPE predictors had a significant association with pCR in multivariate FTV models, there was overall no substantial improvement of the cvAUC (Table 1.3, Model 4). However, we were unable to perform a stratified multivariate analysis within subtype due to limited sample size, which would better mirror the neoadjuvant treatment approach. Our analysis improves on prior literature by evaluating the most relevant clinical question of the additive value of BPE to tumor measurements for predicting pCR, rather than evaluate the utility of BPE prediction alone as most studies do. Changes in the primary tumor are the most direct and robust method for non-invasive prediction of pCR,[23] and the benefit of BPE is therefore most relevant when supplementing tumor models. The only prior study to evaluate the additive effect of BPE for prediction of pCR[22] found that while BPE predictors remained statistically significant in a multiple predictor model, they did not report the extent to which the odds ratio or AUC changed relative to a univariate model.

While BPE still has the potential to be an independent marker of response, our observation of limited additive effect may be due to a variety of reasons. We had a relatively small sample size, which may have caused a strong negative bias when performing cross-validation[24]. There are

also no accepted definitions of quantitative BPE measurement, and thus alternative quantitative techniques (e.g. partial volume sampling or different kinetic parameterizations) should be explored to assess if they have stronger prediction and additive value to FTV models. Finally, given multiple comparisons, our statistically significant univariate BPE results may have been arguably due to chance. However, the fact that we demonstrate a continued improvement in magnitude and strength of BPE prediction with later time points in HR+ cancers indicates a consistent pattern that reduces the likelihood of results being the product of random chance.

CONCLUSION

In conclusion, our results suggest BPE may have subtype specific association with pCR in women with HR+HER2- breast cancer, achieving a similar diagnostic performance to univariate prediction with FTV. However, we did not observe substantial additive improvement in predictive performance when adding BPE to an FTV model in our current study. Additional studies (with ideally larger cohorts) are necessary to replicate these effects, and further understand potentially important additive effects as well as differential effects within subtypes.

REFERENCES

1. *AJCC Cancer Staging Manual*. (Springer, 2017).

2. Symmans WF, Wei C, Gould R, et al. Long-Term Prognostic Risk After Neoadjuvant Chemotherapy Associated With Residual Cancer Burden and Breast Cancer Subtype. J Clin Oncol 2017;35(10):1049-60.

3. Rugo HS, Olopade OI, DeMichele A, et al. Adaptive Randomization of Veliparib–Carboplatin Treatment in Breast Cancer. N Engl J Med 2016;375(1)**:**23–4.

4. Hylton NM, Blume JD, Bernreuter WK, et al. Locally Advanced Breast Cancer: MR Imaging for Prediction of Response to Neoadjuvant Chemotherapy—Results from ACRIN 6657/I-SPY TRIAL. Radiology 2012**;**263(3):663–72.

5. Hylton NM, Gatsonis CA, Rosen MA, et al. Neoadjuvant Chemotherapy for Breast Cancer: Functional Tumor Volume by MR Imaging Predicts Recurrence-free Survival—Results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. *Radiology* 2016;279(1):44-55.

6. Sung JS, Corben AD, Brooks JD, et al. Histopathologic Characteristics of Background Parenchymal Enhancement (BPE) on Breast MRI. Breast Cancer Res Treat 2018;**172(2):**487–96.

7. Kuhl CK, Keulers A, Strobel K, et al. Not All False Positive Diagnoses are Equal: On the Prognostic Implications of False-Positive Diagnoses Made in Breast MRI Versus in Mammography / Digital Tomosynthesis Screening. Breast Cancer Res 2018; 20(1):13.

8. King V, Brooks JD, Bernstein JL, et al. Background Parenchymal Enhancement at Breast MR Imaging and Breast Cancer Risk. Radiology 2011;260(1):50–60 .

9. Dontchos BN, Rahbar H, Partridge SC, et al. Are Qualitative Assessments of Background Parenchymal Enhancement, Amount of Fibroglandular Tissue on MR Images, and Mammographic Density Associated with Breast Cancer Risk? Radiology 2015;276(2): 371–80.

10. Arasu, V. A. *et al.* Population-Based Assessment of the Association Between Magnetic Resonance Imaging Background Parenchymal Enhancement and Future Primary Breast Cancer Risk. J Clin Oncol 2019;37(12):954-63.

11. King V, Goldfarb SB, Brooks JD, et al. Effect of Aromatase Inhibitors on Background Parenchymal Enhancement and Amount of Fibroglandular Tissue at Breast MR Imaging. Radiology 2012;264(3):670–8.

12. Preibsch H, Wanner L, Bahrs SD, et al. Background Parenchymal Enhancement in Breast MRI Before and After Neoadjuvant Chemotherapy: Correlation with Tumour Response. Eur Radiol 2015;26(6):1590-6.

13. Chen JH, Yu H, Lin M, Mehta RS, Su MY. Background Parenchymal Enhancement in the Contralateral Normal Breast of Patients Undergoing Neoadjuvant Chemotherapy Measured by DCE-MRI. Magn Reson Imaging 2013;31(9):1465–71.

14. van der Velden BHM, Dmitriev I, Loo CE, Pijnappel RM, Gilhuijs KGA. Association between Parenchymal Enhancement of the Contralateral Breast in Dynamic Contrast-enhanced MR Imaging and Outcome of Patients with Unilateral Invasive Breast Cancer.

Radiology 2015;276(3):675–85.

15. Jones EF, Sinha SP, Newitt DC, et al. MRI Enhancement in Stromal Tissue Surrounding Breast Tumors: Association with Recurrence Free Survival following Neoadjuvant Chemotherapy. PLoS ONE 2013;8(5): e61969.

16. Grimm LJ, Anderson AL, Baker JA et al. Interobserver Variability Between Breast Imagers Using the Fifth Edition of the BI-RADS MRI Lexicon. AJR Am J Roentgenol 204(5):1120–4.

17. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: An Adaptive Breast Cancer Trial Design in the Setting of Neoadjuvant Chemotherapy. Nature 2009;86(1):1–4.

18. Klifa C, Carballido-Gamio J, Wilmes L, et al. Quantification of Breast Tissue Index from Mr Data Using Fuzzy Clustering. Conf Proc IEEE Eng Med Biol Soc 2004;2004:1667–70.

19. Liao GJ, Henze Bancroft LC, Strigel RM, et al. Background Parenchymal Enhancement on Breast MRI: A Comprehensive Review. J Magn Reson Imaging 2020;51(1):43-61.

20. van der Velden BHM, Sutton EJ, Carbonaro LA, Pijnappel RM, Morris EA, Gilhuijs KGA. Contralateral Parenchymal Enhancement on Dynamic Contrast-Enhanced Mri Reproduces as a Biomarker of Survival in ER-positive/HER2-negative Breast Cancer Patients. Eur Radiol 2018;28(11):4705-16.

21. Chen JH, Yu HJ, Hsu C, Mehta RS, Carpenter PM, Su MY. Background Parenchymal Enhancement of the Contralateral Normal Breast: Association with Tumor Response in

Breast Cancer Patients Receiving Neoadjuvant Chemotherapy. Transl Oncol 2015;8(3):204–9.

22. You C, Peng W, Zhi W, et al. Association Between Background Parenchymal Enhancement and Pathologic Complete Remission Throughout the Neoadjuvant Chemotherapy in Breast Cancer Patients. Transl Oncol 2017;10(5):786–92.

23. Hylton N. MR Imaging for Assessment of Breast Cancer Response to Neoadjuvant Chemotherapy. Magn Reson Imaging Clin N Am 2006;14(3):383–9, vii.

24. Parker BJ, Günter S, Bedo J. Stratification Bias in Low Signal Microarray Studies. BMC Bioinformatics 2007;8:326.

Table 1.1. Participant Characteristics.

	Pathologic complete response (pCR) (N = 29)		Non-pathologic complete response (non-pCR) (N = 59)	
	Mean	**IQR**	**Mean**	**IQR**
Age (years)	46.9	17.0	48.8	12.5
	N	**%**	**N**	**%**
Race				
Asian	3	10%	2	3%
Black or African American	6	21%	7	12%
White	20	69%	50	85%
Menopausal status*				
Pre-menopausal	20	69%	35	59%
Peri/Post-menopausal	9	31%	24	41%
Receptor subtype**				
HR+HER2-	7	24%	36	61%
HR-HER2-	22	76%	23	39%

Abbreviations: IQR, inter-quartile range
Proportions calculated within each column
*There were 20 missing values, which were categorized as pre vs. peri/post-menopausal if age =< 55
**All patients represented were HER2 receptor negative
***All patients received paclitaxel (control) or in combination with an experimental agent for 12 weekly cycles followed by four cycles of anthracycline-cyclophosphamide (AC) every 2-3 weeks prior to surgery

Table 1.2. Univariate analyses of BPE variables for predicting pCR, stratified by HR subtype

Receptor type	ALL		HR+HER2-		HR-HER2-	
Total No. (pCR / non-pCR)	88 (29 / 59)		43 (7 / 36)		45 (22 / 23)	
Predictors	OR (95% CI)	AUC	OR	AUC	OR	AUC
BPE_0	1.02 (0.99-1.05)	0.48	1.04 (0.98-1.10)	0.43	1.02 (0.98-1.07)	0.49
BPE_1	1.00 (0.96-1.04)	0.51	1.02 (0.95-1.10)	0.49	1.00 (0.95-1.07)	0.45
BPE_2	0.96 (0.91-1.00)	0.59	0.95 (0.84-1.03)	0.58	0.96 (0.90-1.02)	0.62
BPE_3	0.95 (0.89-1.01)	0.60	0.88 (0.73-1.00)	0.69	0.97 (0.89-1.05)	0.57
%ΔBPE0_1	0.99 (0.86-1.14)	0.52	0.98 (0.74-1.27)	0.54	1.04 (0.86-1.26)	0.47
%ΔBPE0_2	0.88 (0.75-1.00)	0.60	0.82 (0.58-1.04)	0.67	0.87 (0.69-1.06)	0.59
%ΔBPE0_3	0.87 (0.74-1.00)	0.62	0.64 (0.39-0.92)	0.77	0.91 (0.75-1.09)	0.59

p < 0.05
p < 0.10

Abbreviations: BPE, background parenchymal enhancement; HR, Hormone receptor; pCR, pathologic complete response; non-pCR, non-pathologic complete response; OR, odds ratio; CI, confidence interval; AUC, area under the curve

Nomenclature of predictors: _0, absolute value at pretreatment; _1 or 0_1, absolute value at early treatment or change from baseline to early treatment; _2 or 0_2, inter-regimen or change from baseline to inter-regimen; _3 or 0_3, pre-surgery or change from baseline to pre-surgery

Table 1.3. Comparison of pathologic complete response (pCR) prediction models based on functional tumor volume (FTV) predictors only or adding background parenchymal enhancement (BPE) predictors

Prediction Model	Treatment phase	Predictors	OR (95% CI)	cvAUC
Model 1: Pre-specified FTV variables only	Early treatment	%ΔFTV0_1	0.83 (0.71-0.95)	0.68
		FTV_0	1.00 (0.98-1.01)	
	Inter-regimen	%ΔFTV0_2	0.54 (0.31-0.80)	0.70
		FTV_0	1.00 (0.98-1.01)	
	Pre-surgery	%ΔFTV0_3	0.45 (0.20-0.81)	0.63
		FTV_0	1.00 (0.98-1.01)	
Model 2: Pre-specified BPE & FTV variables only	Early treatment	%ΔFTV0_1	0.89 (0.67-0.93)	0.68
		FTV_0	1.04 (0.98-1.01)	
		%ΔBPE0_1	1.11 (0.94-1.33)	
		BPE_0	1.00 (1.00-1.08)	
	Inter-regimen	%ΔFTV0_2	0.52 (0.28-0.80)	0.68
		FTV_0	1.02 (0.98-1.01)	
		%ΔBPE0_2	0.97 (0.80-1.15)	
		BPE_0	1.00 (0.98-1.07)	
	Pre-surgery	%ΔFTV0_3	0.46 (0.19-0.86)	0.61
		FTV_0	1.01 (0.98-1.01)	
		%ΔBPE0_3	0.94 (0.77-1.13)	
		BPE_0	1.00 (0.97-1.06)	
Model 3: Optimized model using any possible FTV and HR predictors	Any phase of treatment	%ΔFTV0_2	0.52 (0.29-0.78)	0.81
		HR +	0.16 (0.05-0.44)	
Model 4: Optimized model using any possible FTV, HR, BPE predictors	Any phase of treatment	%ΔFTV0_2	0.49 (0.26-0.80)	0.82
		HR +	0.08 (0.02-0.29)	
		BPE_0	1.22 (1.04-1.47)	
		BPE_1	0.83 (0.69-0.98)	
		%ΔBPE0_1	1.93 (1.14-3.53)	
		%ΔBPE0_3	0.86 (0.66-1.06)	

Abbreviations: OR, odds ratio; CI, confidence interval; cvAUC, cross-validated area under the curve (10-repeated 5-fold); HR, hormone receptor

Nomenclature of predictors: _0, absolute value at pretreatment; 0_1, change from baseline to early treatment; 0_2, change from baseline to inter-regimen; 0_3, change from baseline to pre-surgery

$p < 0.05$

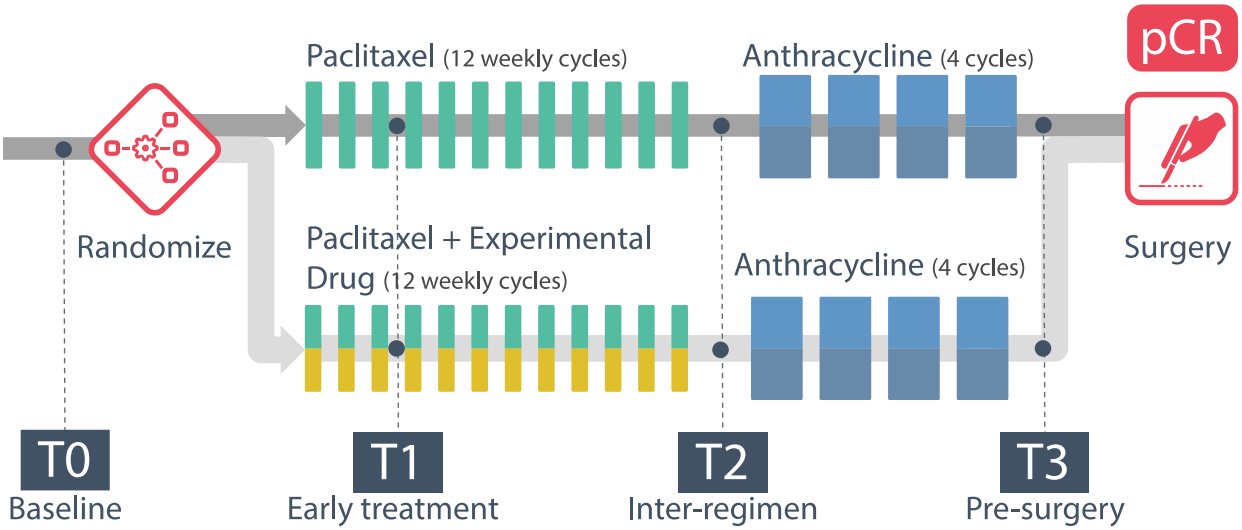


Figure 1.1. I-SPY2 TRIAL schema. I-SPY2 TRIAL study schema and adaptive randomization. Breast MRI was obtained at 4 different time points (T0-T3) as described. Patients were randomized to the control (Paclitaxel) or the experimental drug arm (Paclitaxel + Experimental agent) for 12 weekly cycles followed by four (every 2-3 weeks) cycles of anthracycline-cyclophosphamide (AC) prior to surgery. Pathologic complete response — defined as the absence of residual cancer in the breast or lymph nodes at the time of surgery — is the primary end point of the I-SPY 2 TRIAL.

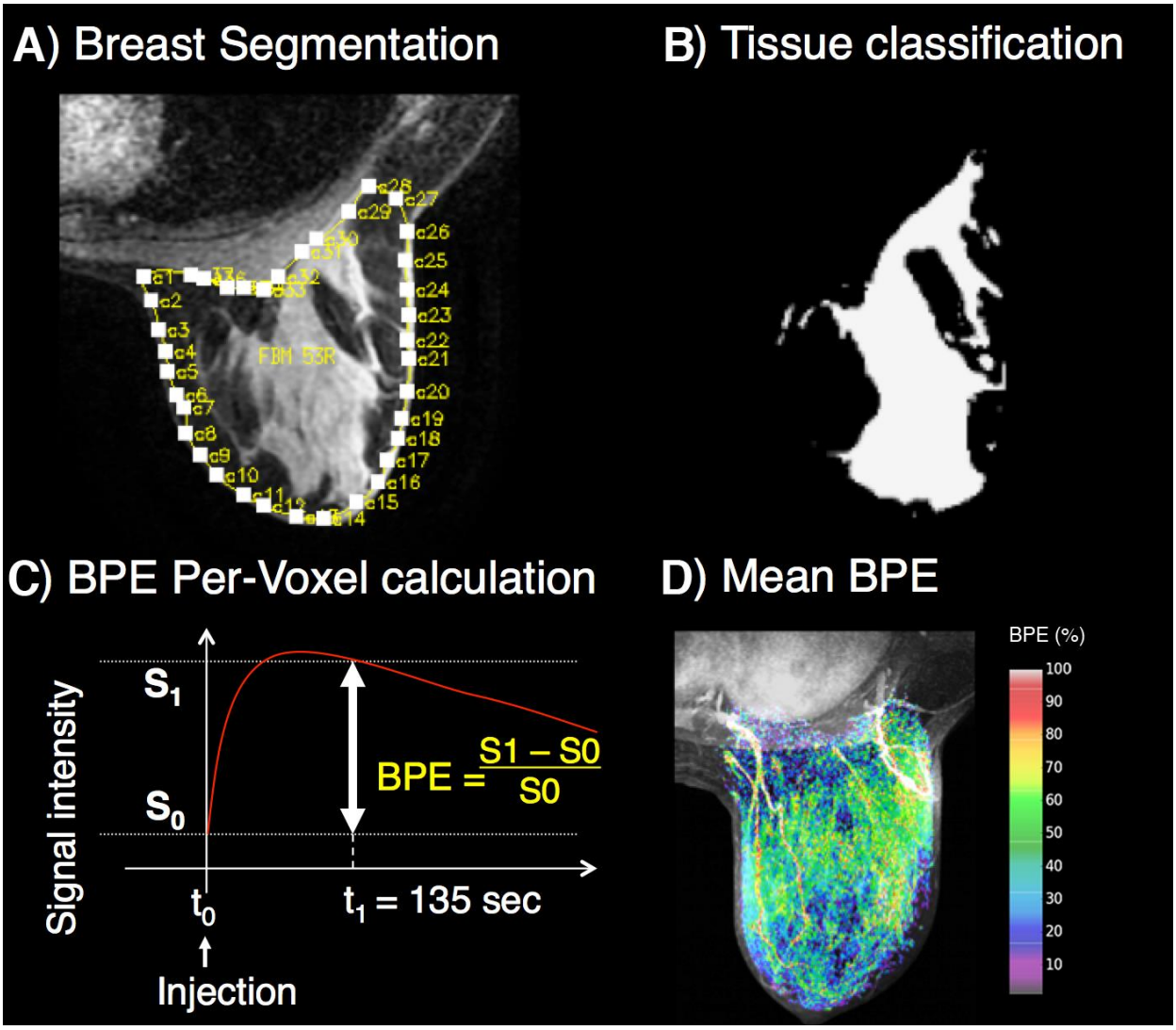


Figure 1.2. Process of quantitative background parenchymal enhancement (BPE) calculation. A: Initially, manual segmentation of the contralateral (unaffected) breast was performed. **B:** This is followed by deriving a mask classifying fibroglandular tissue and removing non-breast elements using fuzzy c-means clustering. BPE is then calculated on per-voxel basis (**C**), and an average value of all voxels is calculated to derive the final BPE estimate (**D**).

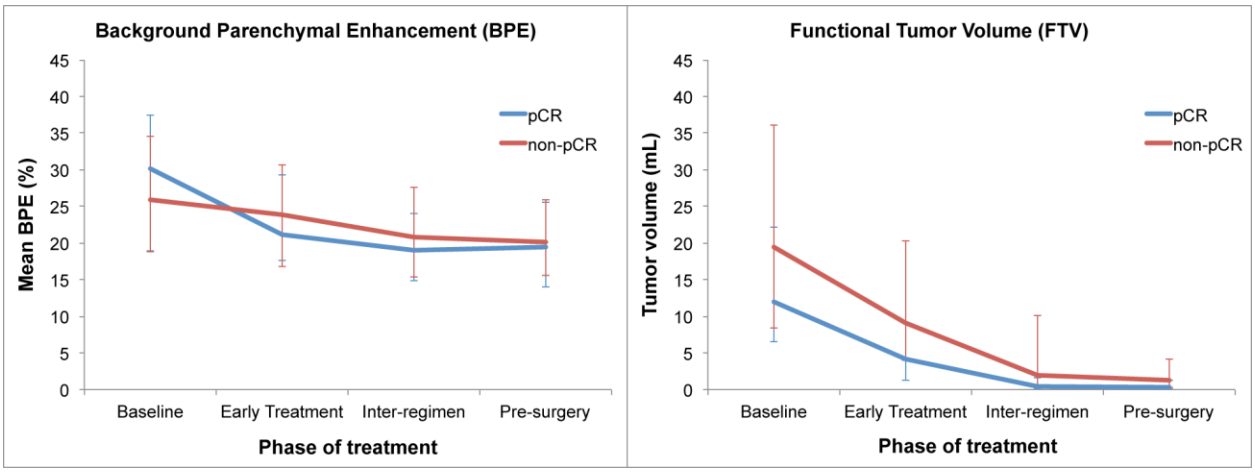


Figure 1.3. Plots of median values of background parenchymal enhancement (BPE) and functional tumor volume (FTV) through phases of treatment (errors bars represent interquartile range). Abbreviation: pCR, pathologic complete response

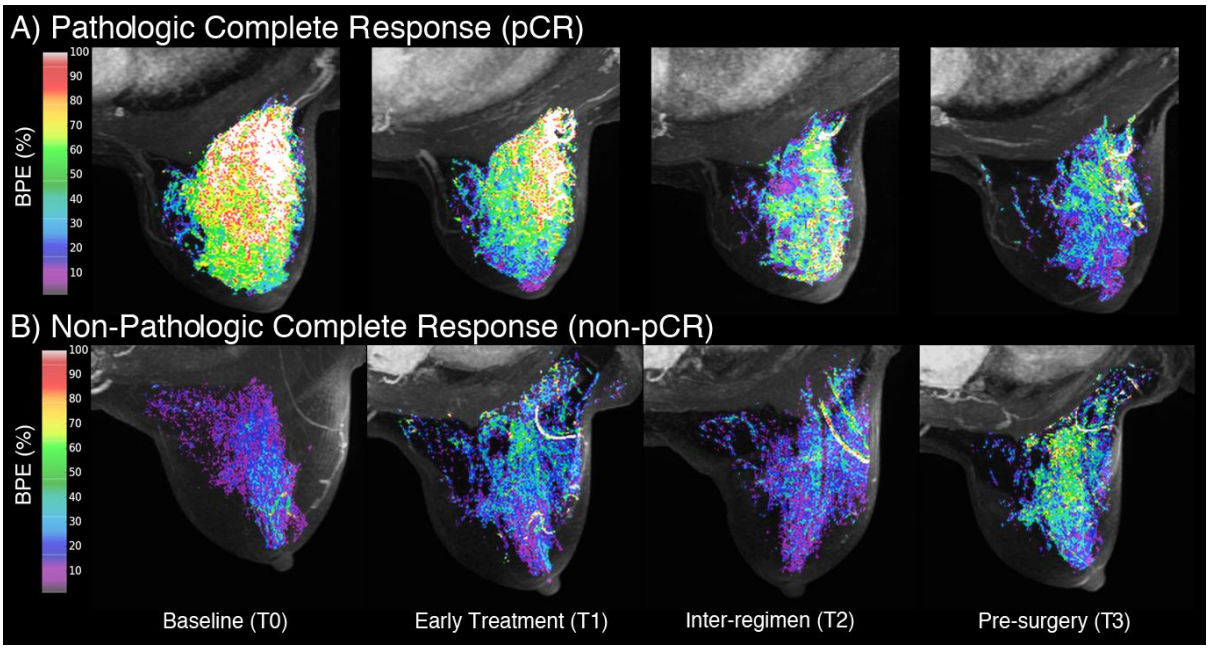


Figure 1.4: Spectral maximum intensity projection breast MRI of an individual woman's background parenchymal enhancement (BPE) at neoadjuvant therapy treatment time points with outcomes of pathologic complete response (pCR) **(A)** or non-pathologic complete response (non-pCR)**(B)**. Women who went on to have pCR (**A**) were more likely to demonstrate higher baseline BPE that decreased with therapy, while women who have non-pCR (**B**) had lower baseline BPE levels that decreased relatively less or did not change with therapy.

CHAPTER 2: Population-Based Assessment of the Association between MRI Background Parenchymal Enhancement and Future Primary Breast Cancer Risk

ABSTRACT

Purpose: To evaluate comparative associations of breast MRI background parenchymal enhancement (BPE) and mammographic breast density with subsequent breast cancer risk.

Patients and Methods: We examined women undergoing breast MRI in the Breast Cancer Surveillance Consortium (BCSC) from 2005-2015 (with one exam in 2000) using qualitative BPE assessments of minimal, mild, moderate, or marked. Breast density was assessed on mammography performed within 5 years of MRI. Among women diagnosed with breast cancer, the first BPE assessment was included if >3 months before their first diagnosis. Breast cancer risk associated with BPE was estimated using Cox proportional hazards regression. Subsequently, an expanded cohort from 2005-2017 was utilized for further subtype evaluations with hormone and Her2-neu tumor receptor in invasive disease, cancer stage, and composite outcome of less favorable cancer defined as Stage IIb or higher, tumor size greater than 15 mm, or positive node status.

Results: Among the initial cohort of 4,247 women, 176 developed breast cancer (129 invasive, 47 ductal carcinoma in situ [DCIS]) over median follow-up of 2.8 years. Mild, moderate, or marked BPE was more common in women with cancer (80%) than in cancer-free women (66%). Compared to minimal BPE, increasing BPE levels were associated with significantly increased cancer risk: mild (hazard ratio (HR)=1.80, 95% confidence interval (CI)=1.12–2.87), moderate (HR=2.42, 95% CI=1.51–3.86) and marked (HR=3.41, 95% CI=2.05–5.66). Compared to women with minimal BPE and almost entirely fatty or scattered fibroglandular breast density,

women with mild, moderate or marked BPE demonstrated elevated cancer risk if they had almost entirely fatty or scattered fibroglandular breast density (HR=2.30, 95% CI=1.19–4.46) or heterogeneous or extremely dense breasts (HR=2.61, 95% CI 1.44–4.72); there was no evidence of interaction between BPE level and breast density (*P*=0.82). Combined mild, moderate or marked BPE was associated with significantly increased risk of invasive cancer (HR=2.73, 95% CI=1.66–4.49), but not DCIS (HR=1.48, 95% CI=0.72–3.05).

The expanded follow-up cohort included 4,944 women, although only 157 total breast cancer cases were included due to withdrawal of data from a single site because of state restrictions on data sharing. For subtype specific evaluations in the expanded cohort, similar significant associations of combined mild, moderate or marked BPE were observed with hormone positive invasive cancer (HR=2.13, 95% CI=1.29—3.54) and early stage (AJCC anatomic stage I/IIa) breast cancer (HR=1.86, 95% CI=1.12—3.10). The magnitude of association increased when using an alternative definition of BPE combining moderate and marked BPE, and statistical significance was also observed for advanced stage (AJCC stage IIb-IV) disease (HR=2.73, 95% CI=1.01—7.38) and less favorable cancer (HR=2.31, 95% CI=1.34—3.97). The association with HER2+ and triple negative disease was not evaluated due to the very limited number of events.

Conclusion: BPE is associated with future invasive breast cancer risk independent of breast density. BPE demonstrates subtype specific associations with less aggressive disease. The association with aggressive subtypes could not be fully evaluated due to sample size limitations, although moderate or marked BPE levels were associated with less favorable disease. BPE should be considered for risk-prediction models for women undergoing breast MRI.

INTRODUCTION

Mammographic breast density is now established as an imaging biomarker for breast cancer risk.[1, 2] Imaging biomarkers are representations of an *in vivo* biological state and phenotype.[3, 4] The incorporation of breast density in breast cancer risk models,[5, 6] as well as state-mandated reporting[7] of mammographic breast density to women, underscores the central role of imaging biomarkers in risk assessment. Recent studies have explored the predictive value of other breast imaging biomarkers, and accumulating evidence suggests elevated background parenchymal enhancement (BPE) assessed on breast magnetic resonance imaging (MRI) may predict primary breast cancer risk.[8-10]

BPE describes the phenomenon observed on breast MRI in which normal breast tissue demonstrates signal enhancement related to uptake of gadolinium-based intravenous contrast, which is used in routine MRI examinations.[3, 4, 11] Biologically, BPE may represent increased tissue microvascularity and/or permeability[5, 6, 12, 13] regulated by endogenous hormones (primarily estrogen),[7, 11] and may represent tissue at risk of neoplasia.[8-10, 14] It is dynamic and variable in appearance and distribution within a woman's breast tissue, and sensitive to the phase of menstrual cycle and lactation,[1, 2, 15] as well as in response to anti-hormonal therapy,[3, 4, 16, 17] chemotherapy,[5, 6, 10, 18-20] and radiotherapy.[7, 21] Similar to mammographic breast density, BPE is qualitatively codified in the Breast Imaging-Reporting and Data System (BI-RADS®) atlas[8-10, 22] as four ordinal levels of increasing enhancement: minimal, mild, moderate, and marked. In contrast to breast density, which is the relative quantity of fat and fibroglandular tissue assessed on mammograms, BPE indicates overall breast tissue contrast enhancement assessed on MRI.

BPE is used clinically to report the level of potential masking of suspicious lesions on MRI, which may impede diagnosis.3, 4, 11, 23, 24 In addition, recent single-center studies have

demonstrated an association between high levels of BPE and increased breast cancer risk.5, 6, 8, 9, 12, 13, 25 In the current study, we evaluate We evaluated the association of BPE and future breast cancer risk among a population-wide cohort of women undergoing breast MRI from diverse practice settings in the US. We compared BPE risk prediction relative to and in conjunction with mammographic breast density. We also subsequently expanded this initial cohort to further investigate subtype specific associations in a follow-up study.

METHODS

Study Setting and Data Sources

We included breast MRIs conducted at 46 radiology facilities that participate in one of six regional Breast Cancer Surveillance Consortium (BCSC) registries (http://www.bcsc-research.org): Carolina Mammography Registry, Kaiser Permanente Washington Registry, Metro Chicago Breast Cancer Registry, New Hampshire Mammography Network, San Francisco Mammography Registry, and Vermont Breast Cancer Surveillance System. BCSC registries link woman-level risk factors and clinical information to breast imaging examinations collected from community radiology facilities. Breast cancer diagnoses and tumor characteristics are obtained by linking with pathology databases; regional Surveillance, Epidemiology, and End Results programs; and state tumor registries.

BCSC registries and the Statistical Coordinating Center received Institutional Review Board approval, and all procedures were Health Insurance Portability and Accountability Act compliant.

Participants and Examinations

We included all BPE measures from screening and diagnostic breast MRI examinations performed on women without a history of breast cancer from 2005–2015 (with one exam in 2000). An expanded cohort from 2005–2017 was further utilized for follow-up study on subtype specific associations. MRI indication was defined as screening or diagnostic by the interpreting radiologist. MRIs were excluded if breast cancer was diagnosed within 3 months following the MRI examination.

Measures and Definitions

For five of the BCSC registries, BPE was assessed clinically as minimal, mild, moderate or marked at the time of MRI interpretation (N=116 radiologists). Although the concept of degrees of parenchymal background enhancement was first published by Kuhl et al. in 2007[12] and BPE was codified formally in American College of Radiology BI-RADS in 2013,[7, 11, 22] awareness and recording of the proposed BI-RADS BPE categories existed before official publication with the first recorded assessments in 2000 in our database. While most BPE assessments were prospectively assessed, a single BCSC registry did not consistently measure BPE clinically. Therefore, a radiologist (N.H.A.) blinded to cancer status retrospectively measured BPE in a subcohort of women with breast cancer and up to 2 matched controls (N=271 MRIs total, of which 38 patients with 52 MRI examinations represented cancer cases). This site had to be withdrawn in the follow-up subtype study due to state laws restricting data sharing.

We primarily dichotomized BPE into minimal vs. mild, moderate or marked BPE based BPE into minimal vs. mild, moderate or marked BPE based on consensus among investigators and prior literature.[8-10, 14] This dichotomized definition was intended to decrease known inter-

reader variability[25-27] for BPE assessment. Breast density was dichotomized using "low density" for almost entirely fatty or scattered fibroglandular densities and "high density" for heterogeneous or extremely dense.

Breast density and risk factors were collected from the closest mammography examination within 5 years of the MRI examination and prior to any breast cancer diagnosis. Women completed a questionnaire at each mammography examination (which were usually performed within 6 months of an MRI) to collect information on race and ethnicity; history of first-degree relatives with breast cancer; menopausal status; and history of breast biopsy. Women were considered postmenopausal if they reported removal of both ovaries, periods that had stopped naturally or had not occurred for more than 365 days, current hormone therapy use, or age 55 or older.[28] Women were considered pre- or peri-menopausal if they reported currently having periods, using oral contraceptives, or not knowing if their periods had stopped.[28] Women were considered to have "surgical menopause, other amenorrhea, or unknown" status if they were under 55 years and reported hysterectomy without bilateral oophorectomy and no use of hormone therapy; reported their periods stopped for "other reasons"; or if menopausal status could not be determined based on available information. Prior diagnoses of benign breast disease were collected from pathology databases and grouped into four categories: non-proliferative, proliferative without atypia, proliferative with atypia, and lobular carcinoma *in situ* as described previously.[29-31] BCSC (version 2.0) 5-year risk score was based on age, race/ethnicity, BI-RADS breast density, first-degree family history, and history of breast biopsy and benign breast disease.[5, 6]

Primary, Secondary and Sensitivity Analysis

We described the participant population at baseline (i.e., first BPE measure) by breast cancer status and BPE. Hazard ratios (HRs) and 95% confidence intervals (CIs) for breast cancer risk were estimated using Cox proportional hazards regression using both ordinal and dichotomized definitions of BPE. We modeled the data in two ways: (1) restricting to each woman's first BPE measure, and (2) including all eligible for each woman. to estimate the impact of multiple measurements. The second model was fit using a robust sandwich estimator for repeated measures survival data to account for multiple observations per woman.[32] Women were followed from 3 months after date of BPE measure to breast cancer diagnosis, death, or end date of complete cancer capture. Models were adjusted for BCSC registry and MRI indication (screening vs. diagnostic), and for number of MRIs in models with multiple measures through stratification. All models were adjusted for age in years as a continuous variable.

BPE was further evaluated in secondary and sensitivity analyses. Associations of BPE with risk were evaluated separately for ductal carcinoma *in situ* (DCIS) and invasive cancer. Multiplicative interaction was tested by including product terms for BPE with breast density, first-degree family history, menopausal status, MRI indication, and BCSC risk score. Confounding was evaluated through adjustment using covariates from Table 2.1. For sensitivity analyses, we refit the model for dichotomous BPE with the following conditions: 1) restricting to breast cancer diagnoses at least one year after BPE measurement and starting follow-up from this time; 2) restricting to BPE measurements assessed in 2010 or later; 3) restricting to non-suspicious BI-RADS assessment categories 1, 2, and 3; and 4) excluding the single registry that retrospectively evaluated BPE.

Follow-up study: Subtype specific exploratory analysis

Similar definitions were used to categorize BPE and breast density, with further exploration using an alternative dichotomized definition of minimal or mild vs. moderate or marked BPE. Subtype specific analyses followed the same baseline and repeated measures models as used in the initial study. The expanded cohort analysis focused on evaluation of associations with hormone and Her2-neu tumor receptor in invasive disease, AJCC version 7 stage, and composite outcome of less favorable cancer defined as Stage IIb or higher, tumor size greater than 15 mm, or positive node status. For a specific subgroup (e.g. invasive cancer), the other competing events (e.g. DCIS) are treated as censored, with the resultant Cox model hazard ratio equivalent to the estimated cause-specific hazard function.

Analyses were performed in SAS® software, version 9.2 (SAS Institute, Cary, NC).) for the initial study and the statistical software R, version 4.0.219 (R Foundation for Statistical Computing, Vienna, Austria), for the follow-up subtype study.

RESULTS

Women and MRI Examination Characteristics

Initial analysis included 6,640 eligible breast MRI examinations conducted in 4,247 women (Table 2.1). Breast MRI examinations were performed for a screening indication in 2,833 women (67%) and a diagnostic indication in 1,414 women (33%). A total of 176 women subsequently developed breast cancer, of whom 129 (73%) had invasive disease and 47 (27%) had DCIS. Median follow-up was 2.1 years for cancer cases (interquartile range = 1.0–3.8 years) and 2.8 years for non-cancer controls (interquartile range =1.4–4.3 years).

Overall, 82% of women were less than 60 years old and 81% were of white/non-Hispanic race and ethnicity (Table 2.1). Women with breast cancer, compared to those without cancer, were slightly more likely to be pre-menopausal (51% vs. 48%), have a first-degree family history of breast cancer (64% vs. 59%), and have a greater ≥1.67% 5-year breast cancer risk by the BCSC model (62% vs. 48%).

When comparing women without breast cancer by BPE group (Table 2.1), women with mild, moderate, or marked BPE compared with minimal BPE were more likely to be less than 60 years old (85% vs. 76%), premenopausal (57% vs. 32%), and have a first-degree family history of breast cancer (61% vs. 56%).

The expanded cohort included 8,167 eligible breast MRI examinations conducted in 4,944 women. Breast MRI examinations were performed for a screening indication in 3,470 women (70%) and a diagnostic indication in 1,474 women (30%). Of the 157 women subsequently developed breast cancer, 114 (73%) had invasive disease and 43 (27%) had DCIS. Although the overall cohort was large, cancer numbers were lower than in the initial study sample due to withdrawal of data from the case control site due to state restrictions on data sharing. Median follow-up was 2.4 years for cancer cases (interquartile range = 1.1–3.8 years) and 3.2 years for non-cancer controls (interquartile range =1.7–5.4 years).

Association of BPE and Cancer

Cancer cases compared to non-cancer cases had a higher proportion of mild, moderate, or marked BPE (80% vs. 66%; Table 2). In the primary analysis using baseline BPE measurement with minimal BPE as reference, increasing levels of BPE demonstrated significantly increased future breast cancer risk: mild (HR=1.80, 95% CI=1.12–2.87), moderate (HR=2.42, 95%

CI=1.51– 3.86), and marked (HR= 3.41, 95% CI=2.05–5.66). Estimates from models using all available BPE assessments for each woman were similar but slightly attenuated. Dichotomous mild, moderate, or marked versus minimal BPE was associated with statistically significantly increased cancer risk in a model using baseline BPE (HR= 2.28, 95% CI=1.51–3.44) or repeated measures of BPE (HR=1.88, 95% CI=1.33–2.65).

Comparative Association of BPE and Breast Density with Cancer

Elevated breast density was more common among women who developed breast cancer (72%) than among women who remained cancer-free (65%; Table 2.3, Figure 2.1). Compared to those with scattered fibroglandular tissue, women with extremely dense breasts demonstrated a non-significant increased risk of breast cancer (HR=1.54, 95% CI=0.97–2.44; Table 2.3), which was had decreased hazard ratio when using repeated measures of breast density.

Both mild, moderate, or marked BPE and high breast density were more common among women who developed breast cancer (57%) than among women who remained cancer-free (38%). Compared to women with low breast density and minimal BPE, women with high breast density and minimal BPE did not have statistically significant increased risk (HR=1.25, 95% CI 0.61-2.54). In contrast, women with low breast density and mild, moderate, or marked BPE had significantly increased risk of breast cancer (HR=2.30, 95% CI 1.19–4.46). Having high breast density with high BPE increased the risk of breast cancer beyond having either factor alone (HR 2.61, 95% CI 1.44–4.72). The test for interaction between dichotomized BPE and dichotomized breast density was non-significant (P=0.82) and therefore in the absence of strong evidence for the interaction we focus on the simpler model without it. Results for the repeated measures model demonstrated similar but attenuated effects.

Comparative Association of BPE and BCSC 5-Year Risk with Cancer

Women with a higher BPE and high BCSC 5-year risk score of ≥1.67% had a four-fold increased hazard compared to women with minimal BPE and low risk score (HR=4.03, 95% CI=1.88–8.63; Table 3). Risk was also elevated but to a lesser extent for higher BPE in the absence of a high risk score (HR=2.91, 95% CI=1.34– 6.31) and marginally elevated for higher risk score in the absence of higher BPE (HR=1.79, 95% CI=0.75–4.30). There was no evidence of a multiplicative interaction between higher risk score and higher BPE on risk (*P*=0.58) and therefore we did not include it. When including multiple examinations, similar statistically significant associations were noted, although with attenuated magnitudes of effect.

Secondary Analyses

Secondary and sensitivity analyses are fully described in Table 4. Mild, moderate, or marked BPE compared to minimal BPE was associated with a statistically significantly increased risk of invasive cancer (HR=2.73, 95% CI=1.66–4.49), but although increased for DCIS (HR=1.48, 95% CI=0.72–3.05) the increase did not reach statistical significance. Higher BPE was associated with an increased risk of breast cancer among women with a first-degree family history (HR=3.55, 95% CI=1.93–6.53). Higher BPE was associated with only slightly increased risk among women without a family history (HR=1.29, 95% CI=0.69–2.42), an association which was not statistically significant. The test for interaction of family history and BPE indicated the difference in the magnitude of the BPE association with breast cancer was unlikely to be attributable to chance (*P* for interaction=0.02). When evaluating confounding through

adjustment, BPE remained significantly associated with risk when adjusting for factors such as breast density, family history, benign breast biopsies, and postmenopausal status.

Sensitivity analyses demonstrated that dichotomous BPE remained significantly associated with cancer risk when restricting to cancer diagnoses made at least 1 year after BPE assessment (HR=2.09, 95% CI=1.34–3.25); restricting to BPE assessments made in or after 2010 (HR=2.99, 95% CI=1.73–5.15); limiting to negative MRI assessments of BI-RADS 1, 2, or 3 (HR=2.14, 95% CI=1.32–3.45); or removing the BCSC registry that retrospectively assessed BPE (HR=2.19, 95% CI=1.43–3.33).

Follow-up study: Subtype specific analysis

In the expanded cohort, associations with overall cancer were similar but slightly attenuated compared to the initial study sample (table 5 and 6). Combined mild, moderate, and marked BPE remained statistically significantly associated with HR+ invasive cancer (HR=2.13, 95% CI=1.29—3.54), HER2- invasive cancer (HR=1.74, 95% CI=1.10—2.74), and early stage I/IIa breast cancer (HR=1.86, 95% CI=1.12—3.10). This association was strengthened when using an alternative dichotomized definition of BPE combining moderate and marked BPE, and statistical significance was also observed for advanced stage IIb-IV disease (HR=2.73, 95% CI=1.01—7.38) and less favorable cancer (HR=2.31, 95% CI=1.34—3.97). The association with HER2+ and triple negative disease was not evaluated due to the limited number of events. The repeated measures model demonstrated the same trends as the baseline BPE model, but with slightly reduced hazard ratio

DISCUSSION

In this population-based assessment of BPE, we demonstrate that among women undergoing screening or diagnostic breast MRI, elevated levels of BPE predict both clinically and statistically significantly higher risk of developing primary invasive breast cancer. BPE had a higher magnitude of hazard association with breast cancer risk than breast density in this population. Moreover, BPE was independent of breast density in risk prediction, and the combination of BPE and breast density increased the overall risk for breast cancer more than either factor alone. Our results strengthen the findings of smaller, single-institution, retrospective studies,[8, 9, 25] and further validate the use of BPE as an imaging biomarker for primary breast cancer risk.

We also demonstrated BPE to have a strong association with invasive cancer, suggesting it is a relevant biomarker for predicting clinically important breast cancer. Furthermore, BPE risk prediction remained significant when adjusting for other factors associated with increased breast cancer risk including increased age, family history, benign breast biopsies, and postmenopausal status. Our population represents a predominantly high-risk group with 49% of women at intermediate-to-high 5-year risk (compared with 38% in a general screened population),[33] which is the primary indication for screening breast MRI.[34] However, in a subset of women of low or average risk (defined by a 5-year BCSC risk score of <1.67%), BPE continued to indicate a significantly increased breast cancer risk. Collectively, these findings suggest that BPE is a robust imaging biomarker for breast cancer risk that is independent of many established factors used in validated risk models.

Our study used the largest longitudinal, population-based cohort to date to confirm the association of BPE with primary breast cancer risk. The validity and robustness of this result is

strengthened by our use of rigorously collected individual-level imaging and pathology data, the majority of which was prospectively obtained from diverse academic and community facilities across the United States. Our findings also validate prior single-institution studies. King et al.[8] initially found moderate or marked BPE was associated with a significantly increased odds for cancer, with an odds ratio of 10.1 (95% CI=2.9–35.3). However, this association may be biased because BPE was measured from MRIs that concurrently displayed enhancing cancer. Dontchos et al.[9] used BPE measurements that preceded cancer diagnosis, and found that mild, moderate or marked BPE was associated with an elevated cancer odds ratio of 9.0 (95% CI=1.1–71.0). All studies including ours found that the significant associations between elevated BPE and breast cancer were greater than associations between breast density and cancer. However, only our study evaluated the interaction between breast density and BPE. While breast density and BPE has been shown to not be correlated among healthy women,[35] we demonstrated that breast cancer risk was independently predicted by breast density and BPE.

In our follow-up study exploring subtype specific associations using an expanded cohort, we continued to show similar patterns of associations of BPE as the initial cohort analysis (table 5 and 6). Additionally, we demonstrated clinically and statistically significant associations with cancers having good prognosis such as HR+ invasive cancer and early stage I/IIa breast cancer. We were limited in our evaluations of more aggressive HER2+ and triple negative subtypes due to the limited sample size in terms of number of events and therefore did not pursue those comparisons. However, using an alternative definition of BPE combining moderate and marked BPE, we demonstrate associations with advanced stage cancer and cancer with less favorable characteristics. Further evaluation with larger cohort size may further elucidate the relationship of BPE to underlying tumor biology and prognosis.

Limitations of using BPE as an imaging biomarker for risk parallel the limitations of breast density. BPE is a qualitative assessment that is prone to interobserver and intraobserver differences that are comparable to or worse than assessment of breast density.[26, 27] BPE has physiological variability, creating sources of measurement error and variation that tend to bias findings towards a null result, which may explain our attenuated results with a repeated measures model. Despite these limitations, BPE was significantly predictive of cancer. Breast density did not significantly predict breast cancer risk despite being an established risk marker; however, this result may be due to higher likelihood of selection of women with dense breast for breast MRI (either for increased individual risk or dense tissue masking), while women with lower breast density selected for breast MRI only for individual risk. To this point, we observe 68% of women in our study had dense breasts, compared to 52% in the general screening population.[36] Importantly, the 95% CI for heterogeneous and extreme density was predominately associated with a clinically significant effect. Finally, BPE prediction remained robust through adjustment for confounders, and sensitivity analyses to remove potential biases related to suspicious assessments on MRI, proximity in time of BPE assessment to cancer diagnosis, and evolving definitions of BPE.

The clinical applicability of BPE as a risk marker is limited to select populations who undergo MRI.[34] Approximately 1–5% of all U.S. women who have received breast imaging have undergone a breast MRI, although this modality may be inappropriately used for some and underutilized for others.[37-39] The indications for and utilization of breast MRI may increase,[40] particularly as it is a potential choice of supplemental screening for women with dense breasts,[41, 42] and/or with recent developments in abbreviated MRI protocols.[43, 44] Information gained with BPE could be helpful for some women in future efforts to better define breast cancer risk and

tailor supplemental screening strategies. For example, if an average-risk woman undergoes diagnostic MRI and demonstrates elevated BPE, her risk may be reassessed to determine if her absolute risk is sufficiently high to warrant screening MRI. Alternatively, high-risk women identified by standard risk-prediction models who undergo screening MRI may demonstrate reduced risk if low BPE levels are considered in conjunction with standard risk models; these women may no longer require routine MRI screening.

In conclusion, we found BPE to be a strong predictor of future breast cancer risk, which was independent of breast density and other established risk factors. BPE demonstrates subtype specific associations with less aggressive disease, evaluation for association with aggressive disease was limited due to sample size and was only clearly noted at moderate and marked levels. BPE should be considered for incorporation into risk-prediction models for women undergoing MRI.

REFERENCES

1. 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. 2015;107(5):dju397.

2. Boyd NF, Guo H, Martin LJ, et al. Mammographic Density and the Risk and Detection of Breast Cancer. N Engl J Med 2007;356:227–36.

3. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology 2015;277:813–25.

4. Hylton N. Dynamic contrast-enhanced magnetic resonance imaging as an imaging biomarker. J Clin Oncol 2006;24:3293–98.

5. Tice JA, Cummings SR, Smith-Bindman R, et al. Using Clinical Factors and Mammographic Breast Density to Estimate Breast Cancer Risk: Development and Validation of a New Predictive Model. Ann Internal Med 2008;148:337–47.

6. Tice JA, Miglioretti DL, Li C-S, et al. Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. J Clin Oncol 2015;33:3137–43.

7. Dehkordy SF, Carlos RC. Dense Breast Legislation in the United States: State of the States. J Am Coll Radiol 2016; 13(11S):R53–7.

8. King V, Brooks JD, Bernstein JL, et al. Background Parenchymal Enhancement at Breast MR Imaging and Breast Cancer Risk. Radiology 2011;260:50–60.

9. Dontchos BN, Rahbar H, Partridge SC, et al. Are Qualitative Assessments of Background

Parenchymal Enhancement, Amount of Fibroglandular Tissue on MR Images, and Mammographic Density Associated with Breast Cancer Risk? Radiology 2015;276:371–80.

10. van der Velden BHM, Dmitriev I, Loo CE, et al. Association between Parenchymal Enhancement of the Contralateral Breast in Dynamic Contrast-enhanced MR Imaging and Outcome of Patients with Unilateral Invasive Breast Cancer. Radiology 2015;276:675–85.

11. Giess CS, Yeh ED, Raza S, et al. Background Parenchymal Enhancement at Breast MR Imaging: Normal Patterns, Diagnostic Challenges, and Potential for False-Positive and False-Negative Interpretation. Radiographics 2014;34:234–47.

12. Kuhl C. The Current Status of Breast MR Imaging. Part I. Choice of Technique, Image Interpretation, Diagnostic Accuracy, and Transfer to Clinical Practice. Radiology 2007;244:356–78.

13. Leung T-K, Huang P-J, Liang H-H, et al. Retrospective Study of False-positive Breast Magnetic Resonance Images and Pathological Results in Taiwan. J Exp Clin Med 2012;4:284–8.

14. Kumar AS, Chen DF, Au A, et al. Biologic Significance of False-Positive Magnetic Resonance Imaging Enhancement in the Setting of Ductal Carcinoma in Situ. Am J Surg 2006;192:520–4.

15. Kuhl CK, Bieling HB, Gieseke J, et al. Healthy Premenopausal Breast Parenchyma in Dynamic Contrast-Enhanced Mr Imaging of the Breast: Normal Contrast Medium Enhancement and Cyclical-Phase Dependency. Radiology 1997;203:137–44.

16. King V, Goldfarb SB, Brooks JD, et al. Effect of Aromatase Inhibitors on Background Parenchymal Enhancement and Amount of Fibroglandular Tissue at Breast MR Imaging. Radiology 2012;264:670–8.

17. Schrading S, Schild H, Kühr M, et al. Effects of Tamoxifen and Aromatase Inhibitors on Breast Tissue Enhancement in Dynamic Contrast–enhanced Breast MR Imaging: A Longitudinal Intraindividual Cohort Study. Radiology 2014;271:45–55.

18. Hattangadi J, Park C, Rembert J, et al. Breast Stromal Enhancement on MRI Is Associated with Response to Neoadjuvant Chemotherapy. Am J Roentgenol 2008;190:1630–6.

19. Jones EF, Sinha SP, Newitt DC, et al. MRI Enhancement in Stromal Tissue Surrounding Breast Tumors: Association with Recurrence Free Survival following Neoadjuvant Chemotherapy. PLoS ONE 2013;8:e61969.

20. Preibsch H, Wanner L, Bahrs SD, et al. Background Parenchymal Enhancement in Breast MRI Before and After Neoadjuvant Chemotherapy: Correlation with Tumour Response. Eur Radiol 2015;26(6):1590-6.

21. Kim YJ, Kim SH, Choi BG, et al. Impact of Radiotherapy on Background Parenchymal Enhancement in Breast Magnetic Resonance Imaging. Asian Pac J Cancer Prev 2014;15:2939–43.

22. D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology, 2013.

23. DeMartini WB, Liu F, Peacock S, et al. Background Parenchymal Enhancement on Breast

MRI: Impact on Diagnostic Performance. Am J Roentgenol 2012;198:W373–80.

24. Ray KM, Kerlikowske K, Lobach IV, et al. Effect of Background Parenchymal Enhancement on Breast MR Imaging Interpretive Performance in Community-based Practices. Radiology 2017; 286(3):822-9.

25. Grimm LJ, Saha A, Ghate SV, et al. Relationship Between Background Parenchymal Enhancement on High-risk Screening MRI and Future Breast Cancer Risk. Acad Radiol 2019; 26(1):69-75.

26. Melsaether A, McDermott M, Gupta D, et al. Inter- and Intrareader Agreement for Categorization of Background Parenchymal Enhancement at Baseline and After Training. American Journal of Roentgenology 2014;203:209–15.

27. Grimm LJ, Anderson AL, Baker JA, et al. Interobserver Variability Between Breast Imagers Using the Fifth Edition of the BI-RADS MRI Lexicon. Am J Roentgenol 2015;204:1120–4.

28. Phipps AI, Ichikawa L, Bowles EJA, et al. Defining Menopausal Status in Epidemiologic Studies: a Comparison of Multiple Approaches and Their Effects on Breast Cancer Rates. Maturitas 2010;67:60–6.

29. Dupont WD, Page DL. Risk Factors for Breast Cancer in Women with Proliferative Breast Disease. N Engl J Med 1985;312:146–51.

30. Page DL, Dupont WD, Rogers LW, et al. Atypical Hyperplastic Lesions of the Female Breast. A Long-Term Follow-Up Study. Cancer 1985;55:2698–708.

31. Page DL, Schuyler PA, Dupont WD, et al. Atypical Lobular Hyperplasia as a Unilateral Predictor of Breast Cancer Risk: A Retrospective Cohort Study. The Lancet 2003;361:125–9.

32. Lee EW, Wei L. Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations. Kluwer Academic Publishers, 1992.

33. Kerlikowske K, Zhu W, Tosteson ANA, et al. Identifying Women With Dense Breasts at High Risk for Interval Cancer. Ann Intern Med 2015;162(10):673–81.

34. Saslow D, Boetes C, Burke W, et al. American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography. CA Cancer J Clin 2007;57(2):75–89.

35. Hansen NL, Kuhl CK, Barabasch A, et al. Does MRI Breast "Density" (Degree of Background Enhancement) Correlate With Mammographic Breast Density? J Magn Reson Imaging 2014;40:483–9.

36. Gierach GL, Ichikawa L, Kerlikowske K, et al. Relationship Between Mammographic Density and Breast Cancer Death in the Breast Cancer Surveillance Consortium. J Natl Cancer Inst 2012;104:1218–27.

37. Wernli KJ, DeMartini WB, Ichikawa L, et al. Patterns of Breast Magnetic Resonance Imaging Use in Community Practice. JAMA Intern Med 2014;174:125–32.

38. Miller JW, Sabatino SA, Thompson TD, et al. Breast MRI Use Uncommon among U.S. Women. Cancer Epidemiol Biomarkers Prev 2013;22(1):159–66.

39. Hill DA, Haas JS, Wellman R, et al: Utilization of Breast Cancer Screening with Magnetic

Resonance Imaging in Community Practice. J Gen Intern Med. 2018;33(3):275-83.

40. Stout NK, Nekhlyudov L, Li L, et al. Rapid Increase in Breast Magnetic Resonance Imaging Use. JAMA Intern Med 2014;174:114–21.

41. Monticciolo DL, Newell MS, Moy L, et al. Breast Cancer Screening in Women at Higher-Than-Average Risk:Recommendations From the ACR. J Am Coll Radiol. 2018;15(3 Pt A):408-14.

42. Horný M, Cohen AB, Duszak R, et al. Dense Breast Notification Laws: Impact on Downstream Imaging After Screening Mammography. Med Care Res Rev 2020;77(2):143-54.

43. Kuhl CK, Schrading S, Strobel K, et al. Abbreviated Breast Magnetic Resonance Imaging (MRI): First Postcontrast Subtracted Images and Maximum-Intensity Projection--A Novel Approach to Breast Cancer Screening With MRI. J Clin Oncol 2014;32:2304–10.

44. Mango VL, Morris EA, Dershaw DD, et al. Abbreviated Protocol for Breast MRI: Are Multiple Sequences Needed for Cancer Detection? Eur J Radiol 2015;84(1):65–70.

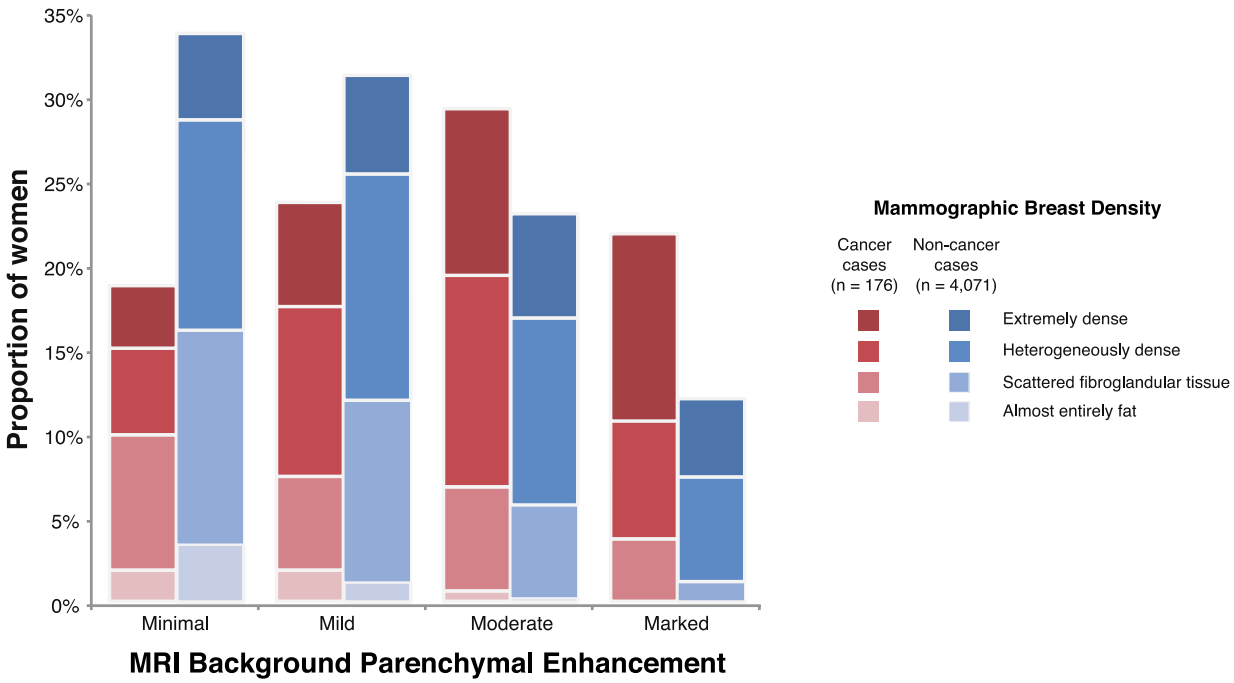


Figure 2.1. Distribution of MRI background parenchymal enhancement (BPE) and mammographic breast density. Red, cancer cases; blue, non-cancer cases. Cancer cases compared to non-cancer cases had a higher proportion of mild, moderate, or marked BPE (80% vs. 66%), and of heterogeneously or extremely dense breasts (72% vs. 65%). When combining BPE and density, cancer cases had higher proportion of both mild, moderate, or marked BPE and heterogeneously or extremely dense breasts (57% vs. 38%).

Table 2.1: Women's characteristics at time of first MRI background parenchymal enhancement (BPE) measurement by cancer status and BPE level.

	Cancer Cases (N=176)			Non-Cancer Cases (N=4,071)		
	Minimal BPE	Mild, moderate, or marked BPE	Total	Mild, moderate, or marked BPE	Minimal BPE	Total
Total women	36	140	176	1,397	2,674	4,071
Age, years						
<40	8 (22)	29 (21)	37 (21)	226 (16)	639 (24)	865 (21)
40-49	5 (14)	51 (36)	56 (32)	361 (26)	986 (37)	1347 (33)
50-59	10 (28)	36 (26)	46 (26)	479 (34)	661 (25)	1140 (28)
60-69	7 (19)	21 (15)	28 (16)	258 (18)	310 (12)	568 (14)
≥ 70	6 (17)	3 (2)	9 (5)	73 (5)	78 (3)	151 (4)
Race/Ethnicity						
White, non-Hispanic	28 (85)	105 (76)	133 (78)	972 (82)	1963 (81)	2935 (81)
Black, non-Hispanic	2 (6)	9 (7)	11 (6)	30 (3)	62 (3)	92 (3)
Hispanic	1 (3)	8 (6)	9 (5)	44 (4)	124 (5)	168 (5)
Asian/Pacific Islander	1 (3)	13 (9)	14 (8)	87 (7)	185 (8)	272 (8)
Other/Mixed	1 (3)	3 (2)	4 (2)	48 (4)	95 (4)	143 (4)
Missing	3 (8)	2 (1)	5 (3)	216 (15)	245 (9)	461 (11)
BI-RADS breast density						
Almost entirely fat	3 (9)	4 (3)	7 (4)	113 (10)	49 (2)	162 (5)
Scattered fibroglandular tissue	13 (39)	25 (19)	38 (23)	419 (38)	576 (26)	995 (30)
Heterogeneously dense	11 (33)	56 (43)	67 (41)	412 (37)	1014 (46)	1426 (43)
Extremely dense	6 (18)	44 (34)	50 (31)	169 (15)	550 (25)	719 (22)
Missing	3 (8)	11 (8)	14 (8)	284 (20)	485 (18)	769 (19)

Table 2.1 (continued)

	Cancer Cases (N=176)			Non-Cancer Cases (N=4,071)		
	Minimal BPE	Mild, moderate, or marked BPE	Total	Mild, moderate, or marked BPE	Minimal BPE	Total
Menopausal status						
Pre/peri-menopausal	7 (26)	66 (57)	73 (51)	316 (32)	1010 (57)	1326 (48)
Post-menopausal	20 (74)	50 (43)	70 (49)	662 (68)	776 (43)	1438 (52)
Surgical or other amenorrhea or unknown	9 (25)	24 (17)	33 (19)	419 (30)	888 (33)	1307 (32)
First-degree family history of breast cancer						
No	15 (50)	42 (32)	57 (36)	483 (44)	915 (40)	1398 (41)
Yes	15 (50)	88 (68)	103 (64)	627 (56)	1385 (60)	2012 (59)
Missing	6 (17)	10 (7)	16 (9)	287 (21)	374 (14)	661 (16)
History of benign breast biopsy						
No	14 (47)	41 (31)	55 (34)	575 (55)	1196 (54)	1771 (55)
Yes	16 (53)	93 (69)	109 (66)	472 (45)	1000 (46)	1472 (45)
Missing	6 (17)	6 (4)	12 (7)	350 (25)	478 (18)	828 (20)
5-year BCSC risk (version 2 calculator)						
<1.00%	5 (19)	21 (18)	26 (18)	227 (23)	489 (25)	716 (24)
1.00–1.66%	4 (15)	26 (22)	30 (21)	277 (28)	538 (27)	815 (27)
1.67–2.49%	9 (33)	25 (21)	34 (23)	251 (25)	477 (24)	728 (25)
2.50–3.99%	7 (26)	37 (31)	44 (30)	179 (18)	329 (17)	508 (17)
≥4.00%	2 (7)	10 (8)	12 (8)	68 (7)	133 (7)	201 (7)
Missing	9 (25)	21 (15)	30 (17)	395 (28)	708 (26)	1103 (27)

Table 2.1 (continued)

	Cancer Cases (N=176)			Non-Cancer Cases (N=4,071)		
	Minimal BPE	Mild, moderate, or marked BPE	Total	Mild, moderate, or marked BPE	Minimal BPE	Total
Cancer Type						
Invasive	25 (69)	104 (74)	129 (73)	–	–	–
DCIS	11 (31)	36 (26)	47 (27)	–	–	–
Indication for MRI at BPE measure						
Screening	22 (61)	82 (59)	104 (59)	951 (68)	1778 (66)	2729 (67)
Diagnostic	14 (39)	58 (41)	72 (41)	446 (32)	896 (34)	1342 (33)
Follow-up interval, median years (interquartile range)	2.1 (1.0, 3.8)		2.8 (1.4, 4.3)		–	

Numbers presented as count (%) unless otherwise indicated. Proportions sum to 100% within column category.
BI-RADS, Breast Imaging-Reporting and Data System; BPE: background parenchymal enhancement; DCIS, ductal carcinoma in situ; MRI, magnetic resonance imaging; BCSC: Breast Cancer Surveillance Consortium

Table 2.2: Associations between baseline MRI background parenchymal enhancement (BPE) and breast cancer risk.

	Baseline BPE			Multiple BPE measures per woman		
	Cancer cases (%)	**Non-cancer cases (%)**	**Hazard ratio* (95% CI)**	**Cancer cases (%)**	**Non-cancer cases (%)**	**Hazard ratio* (95% CI)**
Total	**176 women**	**4,071 women**		**255 observations**	**6,385 observations**	
BI-RADS BPE (4 categories)						
Minimal	36 (20)	1397 (34)	Reference	66 (24)	2342 (37)	Reference
Mild	47 (27)	1257 (31)	**1.80 (1.12-2.87)**	74 (27)	1921 (30)	**1.56 (1.07-2.29)**
Moderate	53 (30)	929 (23)	**2.42 (1.51-3.86)**	71 (26)	1376 (22)	**1.92 (1.28-2.88)**
Marked	40 (23)	488 (12)	**3.41 (2.05-5.66)**	60 (22)	746 (12)	**2.70 (1.68-4.34)**
BPE (dichotomous)						
Minimal	36 (20)	1397 (34)	Reference	66 (24)	2342 (37)	Reference
Mild, moderate, or marked	140 (80)	2674 (66)	**2.28 (1.51-3.44)**	205 (76)	4043 (63)	**1.88 (1.33-2.65)**

Numbers presented as counts (%) unless otherwise indicated. Proportions sum to 100% within column category.

*Hazard ratios estimated from Cox proportional hazards model stratified by BCSC registry and MRI indication, and adjusted for age at BPE measurement. For multiple BPE measures per woman, a robust sandwich estimator was used and models were additionally stratified by MRI number.

BI-RADS, Breast Imaging-Reporting and Data System; BPE: background parenchymal enhancement; CI, confidence intervals; MRI, magnetic resonance imaging

Bolded values represent significant results.

Table 2.3: Associations among mammographic breast density, BCSC 5-year risk, and MRI background parenchymal enhancement (BPE) on breast cancer risk.

	Baseline BPE and/or breast density			Multiple BPE and/or breast density measures per woman		
	Cancer cases (%)	**Non-cancer cases (%)**	**Hazard ratio* (95% CI)**	**Cancer cases (%)**	**Non-cancer cases (%)**	**Hazard ratio* (95% CI)**
Total	**176 women**	**4,071 women**		**255 observations**	**6,385 observations**	
BI-RADS breast density						
Almost entirely fat	7 (4)	162 (5)	1.19 (0.53-2.70)	16 (6)	276 (5)	1.45 (0.64-3.28)
Scattered fibroglandular tissue	38 (23)	995 (30)	Reference	67 (26)	1659 (31)	Reference
Heterogeneously dense	67 (41)	1426 (43)	1.29 (0.85-1.95)	100 (39)	2231 (41)	1.12 (0.77-1.63)
Extremely dense	50 (31)	719 (22)	1.54 (0.97-2.44)	72 (28)	1230 (23)	1.25 (0.81-1.92)
Breast density and BPE**						
Low density + minimal BPE	16 (10)	532 (16)	Reference	33 (13)	930 (17)	
High density + minimal BPE	17 (10)	581 (18)	1.25 (0.61-2.54)	30 (12)	1035 (19)	0.97 (0.59-1.61)
Low density + MMM BPE	29 (18)	625 (19)	**2.30 (1.19-4.46)**	50 (20)	1005 (19)	**1.84 (1.17-2.91)**
High density + MMM BPE	100 (62)	1564 (47)	**2.61 (1.44-4.72)**	142 (56)	2426 (45)	**1.83 (1.24-2.70)**
BI-RADS breast density and BCSC 5-year Risk***						
BCSC risk <1.67% + minimal BPE	9 (6)	505 (17)	Reference	21 (9)	830 (17)	
BCSC risk >=1.67% + minimal BPE	18 (12)	497 (17)	1.79 (0.75-4.30)	33 (14)	966 (20)	1.16 (0.64-2.09)
BCSC risk <1.67% + MMM BPE	48 (33)	1028 (35)	**2.91 (1.34-6.31)**	74 (32)	1507 (31)	**2.02 (1.21-3.36)**
BCSC risk >=1.67% + MMM BPE	71 (49)	938 (32)	**4.03 (1.88-8.63)**	102 (44)	1596 (33)	**2.14 (1.29-3.55)**

Numbers presented as counts (%) unless otherwise indicated. Proportions sum to 100% within column category. Numbers do not sum to total because some women were missing breast density.

*Hazards ratios estimated from Cox proportional hazards model stratified by BCSC registry and MRI indication, and adjusted for age at BPE measurement. For models with multiple BPE measures per woman, a robust sandwich estimator was used and models were additionally stratified by MRI number.

**Test for interaction for baseline BPE and breast density P=0.82; for multiple BPE measures and breast density per woman P=0.99

*** Test for interaction for baseline BPE and BCSC 5-year risk P=0.58; for multiple BPE measures and BCSC 5-year risk per woman P=0.77

Low density: almost entirely fatty or scattered fibroglandular densities; high density: heterogeneously or extremely dense; BCSC: Breast Cancer Surveillance Consortium; BI-RADS, Breast Imaging-Reporting and Data System; BPE: background parenchymal enhancement; CI, confidence intervals; MMM BPE, mild, moderate, or marked BPE

Bolded values represent significant results.

Table 2.4. Secondary and sensitivity analyses of mild, moderate, or marked versus minimal MRI background parenchymal enhancement (BPE) on breast cancer risk.

	Baseline BPE	
	Hazard Ratio* (95% CI)	***P*-value for interaction**
Cancer type		
DCIS	1.48 (0.72-3.05)	
Invasive	**2.73 (1.66-4.49)**	
First-degree family history		
No family history	1.29 (0.69-2.42)	0.02
Family history	**3.55 (1.93-6.53)**	
Menopausal status		
Pre-menopausal	**3.01 (1.27-7.10)**	0.77
Post-menopausal	**2.58 (1.43-4.64)**	
MRI indication		
Screening MRI	**2.58 (1.57-4.25)**	0.22
Diagnostic MRI	1.59 (0.84-2.99)	
Adjusted covariates		
Breast density	**2.22 (1.42-3.47)**	
Family history	**2.34 (1.50-3.65)**	
Benign breast disease (BBD)	**2.21 (1.42-3.43)**	
Family history and BBD	**2.29 (1.45-3.61)**	
Menopausal status	**2.63 (1.56-4.44)**	
Family history, menopausal status, and BBD	**2.53 (1.49-4.28)**	
Sensitivity analyses		
Cancer diagnoses 1 year after BPE only	**2.09 (1.34-3.25)**	
MRI examinations 2010 or later only	**2.99 (1.73-5.15)**	
BI-RADS 1, 2, or 3 exams only	**2.14 (1.32-3.45)**	
Remove site with retrospective BPE assessment	**2.19 (1.43-3.33)**	

*Hazards ratios estimated from Cox proportional hazards model stratified by BCSC registry and MRI indication, and adjusted for age at BPE measurement.

BBD, benign breast disease; BI-RADS, Breast Imaging-Reporting and Data System; BPE: background parenchymal enhancement; CI, confidence intervals; DCIS, ductal carcinoma in situ; MMM BPE, mild, moderate, or marked BPE; MRI, magnetic resonance imaging
Bolded values represent significant results.

Table 2.5a. Cancer subgroup analysis by baseline BPE and/or breast density

Cancer category (sample size)	All Cancers (n = 157)	Invasive Cancer (n = 114)	DCIS (n = 43)
	Hazard ratio* (95% CI)	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**
BI-RADS BPE (4 categories)			
Minimal	Reference	Reference	Reference
Mild	1.23 (0.79 to 1.91)	1.35 (0.79 to 2.29)	0.99 (0.44 to 2.22)
Moderate	**1.9 (1.22 to 2.96)***	**2.35 (1.4 to 3.95)***	1.03 (0.42 to 2.54)
Marked	**2.99 (1.87 to 4.79)***	**3.06 (1.73 to 5.42)***	**2.85 (1.23 to 6.58)***
BPE (dichotomous)			
Minimal	Reference	Reference	Reference
Mild, moderate, or marked	**1.73 (1.2 to 2.49)***	**1.94 (1.25 to 3.02)***	1.29 (0.67 to 2.51)
BPE (dichotomous)			
Minimal or Mild	Reference	Reference	Reference
Moderate or marked	**2.05 (1.48 to 2.84)***	**2.24 (1.53 to 3.28)***	1.62 (0.87 to 3.03)
BI-RADS breast density			
Almost entirely fat	0.98 (0.44 to 2.19)	1.39 (0.61 to 3.18)	Inf
Scattered fibroglandular tissue	Reference	Reference	Reference
Heterogeneously dense	0.89 (0.59 to 1.35)	1.02 (0.63 to 1.65)	0.57 (0.24 to 1.36)
Extremely dense	1.4 (0.9 to 2.18)	1.25 (0.73 to 2.13)	1.88 (0.85 to 4.18)

Table 2.5b. Cancer subgroup analysis by baseline BPE and/or breast density

Cancer category (sample size)	All Cancers (n = 157)	HR+ Invasive (n = 91)	HR- Invasive (n = 20)	HER2+ Invasive (n = 5)	HER2- Invasive (n = 103)	HR-/HER2- Invasive (n = 19)
	Hazard ratio* (95% CI)	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**
BI-RADS BPE (4 categories)						
Minimal	Reference	Reference	Reference	Reference	Reference	Reference
Mild	1.23 (0.79 to 1.91)	1.3 (0.7 to 2.4)	1.37 (0.46 to 4.08)	Inf	1.18 (0.68 to 2.05)	1.18 (0.38 to 3.68)
Moderate	**1.9 (1.22 to 2.96)***	**2.75 (1.53 to 4.95)***	1.28 (0.39 to 4.27)	Inf	**2.29 (1.35 to 3.89)***	1.27 (0.38 to 4.24)
Marked	**2.99 (1.87 to 4.79)***	**3.8 (2.01 to 7.18)***	0.86 (0.17 to 4.32)	Inf	**2.44 (1.32 to 4.5)***	0.85 (0.17 to 4.29)
BPE (dichotomous)						
Minimal	Reference	Reference	Reference	Reference	Reference	Reference
Mild, moderate, or marked	**1.73 (1.2 to 2.49)***	**2.13 (1.29 to 3.54)***	1.24 (0.47 to 3.26)	3.01 (0.47 to 19.24)	**1.74 (1.1 to 2.74)***	1.15 (0.43 to 3.05)
BPE (dichotomous)						
Minimal or Mild	Reference	Reference	Reference	Reference	Reference	Reference
Moderate or marked	**2.05 (1.48 to 2.84)***	**2.73 (1.77 to 4.2)***	0.96 (0.38 to 2.45)	Inf	**2.17 (1.45 to 3.24)***	1.03 (0.4 to 2.66)
BI-RADS breast density						
Scattered fibroglandular tissue	Reference	Reference	Reference	Reference	Reference	Reference
Almost entirely fat	0.98 (0.44 to 2.19)	1.73 (0.7 to 4.32)	0.66 (0.08 to 5.26)	Inf	1.55 (0.67 to 3.57)	0.65 (0.08 to 5.21)
Heterogeneously dense	0.89 (0.59 to 1.35)	1.24 (0.71 to 2.18)	0.51 (0.18 to 1.47)	Inf	1.07 (0.65 to 1.77)	0.5 (0.17 to 1.45)
Extremely dense	1.4 (0.9 to 2.18)	1.78 (0.98 to 3.25)	0.28 (0.06 to 1.34)	5.94 (0.54 to 65.09)	1.19 (0.67 to 2.1)	0.14 (0.02 to 1.11)

Table 2.5c. Cancer subgroup analysis by baseline BPE and/or breast density

Cancer category (sample size)	All Cancers (n = 157)	Stage I/IIa Breast Cancer (n = 86)	Stage IIb/III/IV Breast Cancer (n = 17)	Lymph node Negative (n = 90)	Lymph node Positive (n = 23)	Less favorable cancer** (n = 57)
	Hazard ratio* (95% CI)	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**
BI-RADS BPE (4 categories)						
Minimal	Reference	Reference	Reference	Reference	Reference	Reference
Mild	1.23 (0.79 to 1.91)	1.36 (0.74 to 2.48)	1.48 (0.33 to 6.64)	1.24 (0.7 to 2.19)	0.85 (0.27 to 2.64)	1.01 (0.48 to 2.16)
Moderate	**1.9 (1.22 to 2.96)***	**2.18 (1.2 to 3.99)***	3.67 (0.93 to 14.54)	1.63 (0.9 to 2.97)	2.03 (0.73 to 5.65)	**2.32 (1.16 to 4.62)***
Marked	**2.99 (1.87 to 4.79)***	**2.81 (1.46 to 5.43)***	2.77 (0.54 to 14.19)	**2.8 (1.51 to 5.21)***	0.98 (0.2 to 4.84)	**2.34 (1.03 to 5.32)***
BPE (dichotomous)						
Minimal	Reference	Reference	Reference	Reference	Reference	Reference
Mild, moderate, or marked	**1.73 (1.2 to 2.49)***	**1.86 (1.12 to 3.1)***	2.41 (0.69 to 8.51)	1.62 (1 to 2.61)	1.24 (0.51 to 3.04)	1.65 (0.91 to 2.99)
BPE (dichotomous)						
Minimal, Mild	Reference	Reference	Reference	Reference	Reference	Reference
Moderate, or marked	**2.05 (1.48 to 2.84)***	**2.06 (1.33 to 3.19)***	**2.73 (1.01 to 7.38)***	**1.82 (1.18 to 2.79)***	1.81 (0.76 to 4.29)	**2.31 (1.34 to 3.97)***
BI-RADS breast density						
Almost entirely fat	0.98 (0.44 to 2.19)	1.5 (0.61 to 3.68)	1.13 (0.13 to 9.65)	1.49 (0.56 to 4)	0.75 (0.09 to 6.03)	0.79 (0.18 to 3.45)
Scattered fibroglandular tissue	Reference	Reference	Reference	Reference	Reference	Reference
Heterogeneously dense	0.89 (0.59 to 1.35)	1.14 (0.67 to 1.93)	0.68 (0.2 to 2.37)	1.36 (0.78 to 2.39)	0.36 (0.11 to 1.22)	0.75 (0.36 to 1.53)
Extremely dense	1.4 (0.9 to 2.18)	0.91 (0.48 to 1.74)	0.95 (0.25 to 3.6)	1.79 (0.97 to 3.31)	1.11 (0.39 to 3.14)	1.46 (0.71 to 3)

Numbers presented as counts (%) unless otherwise indicated. Proportions sum to 100% within column category. Numbers do not sum to total because some women were missing breast density.

*Hazards ratios estimated from Cox proportional hazards model stratified by BCSC registry and MRI indication, and adjusted for age at BPE measurement.

**Less favorable cancer: Stage IIB or higher, tumor size greater than 15 mm, or positive node status.

Abbreviations: BCSC: Breast Cancer Surveillance Consortium; BI-RADS, Breast Imaging-Reporting and Data System; BPE: background parenchymal enhancement; CI, confidence intervals; HR, Hormone receptor (estrogen or progesterone); HER2, Her2-neu receptor.

Bolded values represent significant results.

Table 2.6a. Cancer subgroup analysis by multiple BPE and/or multiple breast density measurements

Cancer category (sample size)	All Cancers (n = 243)	Invasive Cancer (n = 187)	DCIS (n = 56)
	Hazard ratio* (95% CI)	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**
BI-RADS BPE (4 categories)			
Minimal	Reference	Reference	Reference
Mild	1.25 (0.84 to 1.86)	1.39 (0.87 to 2.21)	0.88 (0.44 to 1.8)
Moderate	**1.57 (1.02 to 2.44)***	**1.79 (1.13 to 2.85)***	1.02 (0.34 to 3.08)
Marked	**2.62 (1.63 to 4.21)***	**2.68 (1.55 to 4.63)***	2.41 (0.92 to 6.29)
BPE (dichotomous)			
Minimal	Reference	Reference	Reference
Mild, moderate, or marked	**1.59 (1.12 to 2.24)***	**1.74 (1.18 to 2.56)***	1.19 (0.58 to 2.43)
BPE (dichotomous)			
Minimal or Mild	Reference	Reference	Reference
Moderate or marked	**1.74 (1.23 to 2.47)***	**1.79 (1.21 to 2.64)***	1.59 (0.72 to 3.51)
BI-RADS breast density			
Almost entirely fat	1.64 (0.8 to 3.38)	**2.18 (1.02 to 4.66)***	0.3 (0.04 to 2.09)
Scattered fibroglandular tissue	Reference	Reference	Reference
Heterogeneously dense	0.83 (0.54 to 1.27)	0.99 (0.62 to 1.58)	0.45 (0.18 to 1.12)
Extremely dense	1.23 (0.78 to 1.94)	1.17 (0.69 to 1.98)	1.54 (0.65 to 3.65)

Table 2.6b. Cancer subgroup analysis by multiple BPE and/or multiple breast density measurements

Cancer category (sample size)	All Cancers (n = 157)	HR+ Invasive (n = 138)	HR- Invasive (n = 36)	HER2+ Invasive (n = 7)	HER2- Invasive (n = 36)	HR-/HER2- Invasive (n = 35)
	Hazard ratio* (95% CI)	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**
BI-RADS BPE (4 categories)						
Minimal	Reference	Reference	Reference	Reference	Reference	Reference
Mild	1.23 (0.79 to 1.91)	1.68 (0.98 to 2.89)	1.13 (0.43 to 2.97)	Inf	1.29 (0.81 to 2.06)	1.05 (0.38 to 2.87)
Moderate	**1.9 (1.22 to 2.96)***	**2.61 (1.55 to 4.41)***	0.75 (0.27 to 2.04)	2.09 (0.51 to 8.64)	**1.86 (1.17 to 2.96)***	0.75 (0.27 to 2.05)
Marked	**2.99 (1.87 to 4.79)***	**4.13 (2.26 to 7.55)***	0.64 (0.12 to 3.25)	Inf	**2.28 (1.31 to 3.97)***	0.64 (0.12 to 3.28)
BPE (dichotomous)						
Minimal	Reference	Reference	Reference	Reference	Reference	Reference
Mild, moderate, or marked	**1.73 (1.2 to 2.49)***	**2.36 (1.48 to 3.75)***	0.9 (0.43 to 1.89)	8.62 (0.49 to 150.7)	**1.64 (1.12 to 2.4)***	0.87 (0.41 to 1.83)
BPE (dichotomous)						
Minimal or Mild	Reference	Reference	Reference	Reference	Reference	Reference
Moderate or marked	**2.05 (1.48 to 2.84)***	**2.39 (1.61 to 3.55)***	0.67 (0.22 to 2.03)	Inf	**1.77 (1.19 to 2.62)***	0.69 (0.22 to 2.15)
BI-RADS breast density						
Almost entirely fat	0.98 (0.44 to 2.19)	2.05 (0.78 to 5.38)	2.53 (0.65 to 9.81)	0	**2.36 (1.08 to 5.17)***	2.52 (0.65 to 9.79)
Scattered fibroglandular tissue	Reference	Reference	Reference	Reference	Reference	Reference
Heterogeneously dense	0.89 (0.59 to 1.35)	1.08 (0.64 to 1.84)	0.58 (0.19 to 1.73)	0	0.94 (0.58 to 1.53)	0.58 (0.2 to 1.74)
Extremely dense	1.4 (0.9 to 2.18)	1.57 (0.89 to 2.8)	0.39 (0.07 to 2.03)	4.09 (0.63 to 26.74)	1.1 (0.62 to 1.95)	0.29 (0.04 to 2.18)

Table 2.6c. Cancer subgroup analysis by multiple BPE and/or multiple breast density measurements (cont)

Cancer category (sample size)	All Cancers (n = 243)	Stage I/IIa Breast Cancer (n = 136)	Stage IIb/III/IV Breast Cancer (n = 23)	Lymph node Negative (n = 132)	Lymph node Positive (n = 43)	Less favorable cancer** (n = 89)
	Hazard ratio* (95% CI)	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**	**Hazard ratio* (95% CI)**
BI-RADS BPE (4 categories)						
Minimal	Reference	Reference	Reference	Reference	Reference	Reference
Mild	1.25 (0.84 to 1.86)	1.63 (0.97 to 2.74)	1.26 (0.32 to 4.9)	1.34 (0.82 to 2.19)	0.58 (0.21 to 1.57)	0.87 (0.45 to 1.71)
Moderate	**1.57 (1.02 to 2.44)***	**1.94 (1.15 to 3.27)***	**3.66 (1.33 to 10.02)***	1.45 (0.83 to 2.55)	1.42 (0.6 to 3.38)	1.8 (0.97 to 3.36)
Marked	**2.62 (1.63 to 4.21)***	**2.48 (1.36 to 4.51)***	3.68 (0.77 to 17.55)	**2.42 (1.35 to 4.35)***	1.17 (0.26 to 5.32)	**2.61 (1.16 to 5.9)***
BPE (dichotomous)						
Minimal	Reference	Reference	Reference	Reference	Reference	Reference
Mild, moderate, or marked	**1.59 (1.12 to 2.24)***	**1.88 (1.23 to 2.87)***	2.45 (0.85 to 7.08)	**1.56 (1.02 to 2.4)***	0.93 (0.41 to 2.13)	1.45 (0.82 to 2.57)
BPE (dichotomous)						
Minimal, Mild	Reference	Reference	Reference	Reference	Reference	Reference
Moderate, or marked	**1.74 (1.23 to 2.47)***	**1.65 (1.05 to 2.59)***	**3.32 (1.35 to 8.13)***	**1.55 (1.02 to 2.36)***	1.62 (0.72 to 3.65)	**2.19 (1.33 to 3.61)***
BI-RADS breast density						
Almost entirely fat	1.64 (0.8 to 3.38)	2.06 (0.83 to 5.13)	1.07 (0.12 to 9.9)	**2.64 (1.12 to 6.22)***	2.43 (0.61 to 9.69)	1.49 (0.47 to 4.77)
Scattered fibroglandular tissue	Reference	Reference	Reference	Reference	Reference	Reference
Heterogeneously dense	0.83 (0.54 to 1.27)	1.06 (0.64 to 1.75)	0.87 (0.19 to 3.97)	1.35 (0.82 to 2.23)	0.59 (0.16 to 2.12)	0.89 (0.4 to 1.96)
Extremely dense	1.23 (0.78 to 1.94)	0.93 (0.5 to 1.71)	1.26 (0.35 to 4.53)	1.49 (0.84 to 2.64)	1 (0.33 to 3)	1.77 (0.85 to 3.66)

Numbers presented as counts (%) unless otherwise indicated. Proportions sum to 100% within column category. Numbers do not sum to total because some women were missing breast density.

*Hazards ratios for models with multiple BPE measures per woman were estimated from Cox proportional hazards model stratified by BCSC registry, MRI indication, and MRI number, and adjusted for age at BPE measurement. A robust sandwich estimator was used for estimating standard errors.

**Less favorable cancer: Stage IIB or higher, tumor size greater than 15 mm, or positive node status.

Abbreviations: BCSC: Breast Cancer Surveillance Consortium; BI-RADS, Breast Imaging-Reporting and Data System; BPE: background parenchymal enhancement; CI, confidence intervals; HR, Hormone receptor (estrogen or progesterone); HER2, Her2-neu receptor.

Bolded values represent significant results.

CHAPTER 3: Comparison of Mammography Artificial Intelligence Algorithms for 5-year Breast Cancer Risk Prediction

ABSTRACT

Importance: Predicting future risk of breast cancer can inform screening and prevention strategies.

Objective: To examine whether mammography artificial intelligence (AI)–based computer vision algorithms, most of which are trained to detect visible breast cancer, can also predict future risk.

Design: Case-cohort study.

Setting: Kaiser Permanente Northern California.

Participants: Women who had a screening mammogram with no evidence of cancer on final imaging assessment in 2016, were followed until 2021. Women with prior breast cancer or highly penetrant gene mutation were excluded. A random subcohort of 13,881 (4.2%) was selected from 329,814 eligible women, of whom 197 had incident cancer. All 4,475 additional incident cases among eligible women were also included.

Exposure: Five available AI algorithms were compared with the clinical risk model developed by the Breast Cancer Surveillance Consortium (BCSC v2). Continuous AI scores were generated from algorithms using the index 2016 mammogram.

Main Outcome and Measures: Risk estimates for the main outcome, incident breast cancer within 0 to 5 years after the index mammogram, were generated using the Kaplan-Meier method and time-varying area under the curve [AUC(t)].

Results: For incident cancers at 0-1 year (interval cancer risk), BCSC demonstrated an AUC(t) of 0.62 (95%CI, 0.58-0.66), and AI algorithms had AUC(t)s of 0.66-0.71 that were all statistically significantly higher ($p < 0.05$). For incident cancers at 1 to 5 years (5-year future cancer risk), BCSC demonstrated an AUC(t) of 0.61 (95% CI, 0.60-0.62), and all AI algorithms had significantly higher AUC(t)s of 0.63-0.67. Combined models using BCSC and AI gave AUC(t)s for interval cancer risk of 0.67-0.73 that was significantly higher for 2 of 5 models compared to AI alone, and gave 5-year future cancer risk of 0.65-0.68 that were significantly higher for all models.

Conclusion and Relevance: All mammography AI algorithms had significantly higher discrimination than the BCSC clinical risk model for interval cancer and 5-year future cancer risk. Combined AI and BCSC models significantly improved performance than AI alone for most comparisons and should be considered for incorporation in future breast cancer models.

INTRODUCTION

Breast cancer risk prediction models are used to evaluate and guide a range of clinical considerations, including hereditary risk, supplemental screening, and risk-reducing medications.[1] Risk models are also under active investigation for broader population management, such as risk-based personalized screening[2,3] or capacity management.[4] Several models have been developed to assess the risk for breast cancer in the general population, including Breast Cancer Risk Assessment Tool (BCRAT, also known as Gail),[5] Breast Cancer Surveillance Consortium (BCSC),[6,7] and International Breast Cancer Intervention Study (also known as Tyrer-Cuzick).[8] Beyond age, these models include clinical factors (eg, family history of breast cancer, race/ethnicity, prior benign breast biopsy), genetic factors, and mammographic breast density. However, these models have moderate discrimination for predicting either 5 or 10-year risk of breast cancer, with areas under the curve (AUC) ranging from 0.62 to 0.66. Computer vision–based artificial intelligence (AI) models have the potential to improve risk prediction beyond clinical risk factors. These models quantitatively extract imaging biomarkers that represent underlying pathophysiologic mechanisms and phenotypes.[9] Breast density is a single imaging biomarker most commonly incorporated into clinical risk models, but recent advances in AI deep-learning[10] provide the ability to extract hundreds to thousands of additional mammographic features beyond breast density alone. However, most mammography-based AI algorithms have been explicitly trained to assist a radiologist with detecting cancer visible on screening mammography (i.e. a short time horizon) and not trained to help predict future risk several years after the time of examination.[11] A few studies have evaluated future risk performance for AI algorithms explicitly trained for this task; these studies suggest substantial improvements over clinical risk models alone.[12,13] However, it is unknown if AI trained for

detection at shorter time horizons, which comprise the majority of mammography AI algorithms and primarily used in clinical practice, also carries longer term predictive performance.

We evaluated 5 commercial and academic mammography AI algorithms, variously trained for short to long time horizons, for the predictive performance of a 5-year time horizon following a negative mammogram in a large, community-based US cohort from the Kaiser Permanente Northern California (KPNC) integrated health system. We compared AI performance to the well-validated BCSC clinical risk model and explored whether combining AI and BCSC clinical risk models can further improve risk prediction above either model type alone.

Methods

Study Design, Setting, and Population

We performed a retrospective case-cohort study of women who had a bilateral screening mammogram in 2016 at KPNC (ie, index mammogram), without evidence of cancer on final imaging assessment either at the time of screening are after diagnostic work-up of positive screening findings. Women were excluded if they had a prior history of breast cancer or a high-penetrance breast cancer susceptibility gene as defined by the National Comprehensive Cancer Network guidelines.[14] Of 329,814 women who met these criteria, a random subcohort of 13,881 women (4.2%) containing 197 cases were selected for analyses, plus all 4,475 additional incident cases diagnosed within 5 years of the index 2016 mammogram (4,672 total or 100% of cases; Figure 1). This sample size was based on the maximum cohort size feasible for AI algorithm evaluation. Our study was reviewed and approved by our regional institutional review board for HIPAA compliance. The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) and case-cohort–specific reporting guidelines.[15,16]

Primary Outcome Ascertainment: Breast Cancer

The primary outcome was incident breast cancer defined as pathologically confirmed invasive carcinoma or ductal carcinoma in situ (DCIS). Cancers were ascertained from the KPNC Breast Cancer Tracking System[17] quality assurance program, which has a 99.8% concordance with the KPNC tumor registry that reports to the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program, but the KPNC Breast Cancer Tracking System identifies incident cancers more rapidly (within 1 month of diagnosis) with manual verification. Women were followed from their index mammogram to date of breast cancer diagnosis; death; health plan disenrollment (allowing up to a 3-month gap in health plan enrollment); or August 31, 2021, whichever occurred first.

Primary Predictor Data Source: Negative Screening Mammograms from 2016

Screening mammograms in 2016 were identified by a Current Procedural Terminology examination code of 77057. A screening mammogram was considered to have no evidence of cancer on final imaging assessment based on Breast Imaging Reporting and Data System (BI-RADS) assessment category 1 or 2 on the screening mammogram, BI-RADS 0 on the screening mammogram and BI-RADS 1 or 2 on the diagnostic mammogram within 90 days, or BI-RADS 0 on the screening mammogram and BI-RADS 4 or 5 on the diagnostic mammogram with concordant benign biopsy within 90 days. The mammograms were evaluated in their archived processed form and were predominately acquired on Hologic stations (87%), followed by General Electric (13%). In the KPNC health system, most average-risk women start screening mammography at age 50 with a screening frequency of every 2 years, although women are given the option to screen starting at age 40 or to screen annually.

Primary Predictor Measurement: AI Risk Score Derived from Screening Mammogram

AI scores were generated from 5 academic and commercial deep-learning computer vision algorithms that take the screening mammogram images as their input, and output patient-level predicted scores. Candidate algorithms were chosen from an ongoing institutional AI operational evaluation. We evaluated 2 academic algorithms freely available for research, the Mirai algorithm (Massachusetts Institute of Technology, Boston, Massachusetts; https://www.github.com/yala/Mirai)[13] and the Globally Aware Multiple Instance Classifier (GMIC) algorithm (New York University, New York City, New York; https://www.github.com/nyukat/gmic). The 3 commercial vendor identities were anonymized due to confidentiality, and labeled Vendor A, Vendor B, and Vendor D. All algorithms were trained for time horizons between 3 months to 5 years, but we evaluated the extent to which all algorithm's predicted score can predict future risk up to 5 years. Further details of the software architecture for each algorithm are provided in the supplement. Because the Mirai algorithm was also calibrated to provide predicted absolute risk scores, further evaluation of calibration was performed for the Mirai algorithm and the BCSC risk score. When any algorithm failed to process an individual mammogram, this missing score was imputed using the algorithm's specific overall median score (missingness by algorithm is detailed in Table 3.1; evaluation by scored cases only is provided in Table 3.2).

Comparative Predictor Ascertainment: Clinical Risk Score

The BCSC clinical 5-year risk prediction model version 2[6,18] was used as the comparative predictor to the AI models. The BCSC model predicts risk for women without a history of breast cancer based on age, race/ethnicity, first-degree family history of breast cancer, prior benign

breast biopsy, and mammographic breast density. Risk factors were obtained in the KPNC electronic health record for risk score generation. Clinical risk factor data were obtained at or before the index date of the first screening mammogram in 2016, regardless of prior membership in the KPNC health system to reflect the underlying population. Breast density was based on the index mammogram using the BI-RADS classification system and from assessments done as part of routine care by radiologists at the time of interpretation. Data completeness for family history and prior history of breast cancer was dependent on patient responses to clinic intake forms or recorded by the provider during routine care; our data structure does not distinguish a woman with no family history from a woman with missing data. Breast biopsy data were available if obtained while the woman was enrolled in the KPNC health plan. Although our biopsy database prospectively classifies atypia and lobular carcinoma in situ, it does not distinguish proliferative benign pathology from otherwise benign pathology, so these benign outcomes were conservatively classified as non-proliferative lesions. Although the Mirai algorithm can input clinical variables for a combined risk score, at the time of evaluation this feature was not useable because it did not allow a variable amount of missing risk factors.

Statistical Analysis

The statistical software R, version 4.0.2[19], was used for all statistical analyses. All statistical tests were 2-sided, with the level for statistical significance set at $\alpha = .05$. We evaluated the ability of the measures to predict breast cancer occurring within three time periods following mammography: "interval cancer risk" as incident cancers at diagnosed between 0 to 1 years, "future cancer risk" as incident cancers diagnosed between 1 to 5 years, and "overall 5-year risk" as incident cancers diagnosed between 0 to 5 years.

Kaplan-Meier estimators were used to estimate the overall 5-year cumulative incidence of breast cancer within strata of each risk score (lowest percentile, middle 80%, highest percentile); design weights were included to account for the case-cohort sampling. Discrimination was evaluated through the time-dependent area under the curve [AUC(t)], which accounts for the dynamic definition of cases and non-cases when handling time-to-event outcomes,[20] for time horizons of 1, 2, 3, 4, and 5 years. We implemented the estimator that accounts for the censoring and sampling distribution using inverse probability of censoring weights and the case-cohort sampling[21] and obtained corresponding 95% confidence intervals (CIs) using bootstrapping with 1000 bootstrap samples.[22] To compare AUC(t) estimates from 2 separate risk scores (eg, BCSC vs. Mirai), we calculated the difference in estimates and corresponding bootstrapped 95% CIs; a CI that does not contain 0 indicates that the difference in AUC(t) estimates is statistically significant at the .05 level.[23]

A Cox model was fit to predict 5-year risk by using the combined screening mammogram–derived AI predicted score and the clinical BCSC risk score. The Cox models accounted for the case-cohort sampling through design weights and included both the AI score and clinical risk score flexibly using restricted cubic splines with 4 knots.[24,25] We used 5-fold cross-validation (CV) to estimate the AUC(t) estimator described above[21] and present the average value across the 5-folds. We obtained corresponding 95% CIs for the average CV-AUC(t) through bootstrapping with 1000 bootstrap samples.

Calibration was estimated for just the Mirai algorithm and BCSC clinical risk model because these are the only models that generate absolute risk estimates. We assessed the calibration of risk scores within prespecified strata of 5-year risk (0 to <1.67%, 1.67 to <3%, ≥3%) based on thresholds established by the BCSC. We compared the observed number of cases

over the 5-year study period with the expected number of cases, calculated as the sum of the cumulative hazard estimates over all individuals in the study.[26] We report the ratio of observed to expected cases with exact 95% CIs.[27] We calculated the incidence rates (IRs; cases per 1000 person-years) and IR ratios (IRRs) with 95% CIs based on a Poisson distribution.[28] All expected incidence estimates incorporated design weights that account for the case-cohort sampling.

Results

Patient Characteristics

The overall baseline patient characteristics are described in Table 3.3. Although our population was predominately older and non-Hispanic white women, we had a substantial proportion of women who were younger than age 50 (23%) or of non-White race/ethnicity (48%). Women had 5.0 median years of follow-up (interquartile range 4.7 to 5.3 years). Women were censored due to end of follow-up (92%), disenrollment (6%), or death (2%).

Cumulative Incidence Rates of BCSC Clinical Risk Model and AI algorithm scores

Figure 2 describes the cumulative incidence rates for breast cancer over 5 years by each risk model. The BCSC average cumulative IR at 5 years in women with >90% percentile of risk was 30.3 per 1000 (95% CI, 28.0-32.9 per 1000), with middle 80% risk the IR was 15.0 per 1000 (95% CI, 14.5-15.7 per 1000), and with <10% percentile of risk the IR was 6.1 per 1000 (95% CI, 5.1-7.2 per 1000). The IRR of the highest to lowest percentile of risk was 5.4. Women with a >90% percentile BCSC risk predicted 21% of all cancers by 5 years, while women with <10% percentile risk predicted 3% of all cancers.

For AI algorithms, the average cumulative IR at 5 years in women with >90% percentile risk ranged from 31.3 to 40.9 per 1000; in women with middle 80% percentile risk, from 13.7 to 15.0 per 1000; and with <10% percentile risk from 6.4 to 7.4 per 1000. The IRR of the highest to lowest percentile of risk ranged between 5.3 and 7.3. Women with >90% percentile AI risk predicted 20% to 27% of all cancers by 5 years, whereas women with <10% percentile risk predicted approximately 2% to 4% of cancers across all AI algorithms.

Discrimination and Calibration of BCSC Clinical Risk Model and AI algorithm scores

When evaluating discrimination (Table 3.4) for incident cancers at 0-1 year (interval cancer risk), BCSC demonstrated an AUC(t) of 0.62 (95% CI, 0.58-0.66). In comparison, all AI algorithms demonstrated an interval cancer risk between 0.66 and 0.71 that was statistically significantly higher than BCSC ($P < .05$). For incident cancers at 1 to 5 years (5-year future cancer risk), BCSC demonstrated an AUC(t) of 0.61 (95% CI, 0.60-0.62). In comparison, 5-year future cancer risk AI algorithms ranged from 0.61 to 0.67 with all algorithms statistically significantly higher than BCSC ($P < .05$).

When evaluating combined AI and BCSC risk models, we initially created linear combination of these predictors. However, all models violated the proportional hazards assumption and were therefore modeled as restricted cubic splines. Combined models demonstrated an AUC(t) for interval cancer risk that ranged from 0.67 to 0.73 (Table 3.5), that was only significantly higher than the corresponding AI algorithm alone for Mirai and GMIC. The combined model 5-year future cancer risk AUC(t) ranged from 0.65 to 0.68 and was significantly higher for all models.

Additional subgroup analyses (supplementary Tables 3.2, 3.6–3.10) demonstrate comparable discrimination to Table 3.4 when restricting to women with invasive breast cancer only, complete scores available across all models only, BI-RADS 1 or 2 on screening mammograms only, and mammograms acquired on Hologic equipment only. Discrimination was mixed for women with BI-RADS 0 on screening mammograms only and mammograms acquired on GE equipment only.

Comparing calibration of the BCSC risk model and the Mirai algorithm (Table 3.11), the 5-year calibration of the BCSC ranged from 1.02 to 1.07 depending on the prespecified risk threshold ranges, whereas that of the Mirai algorithm ranged from 0.49 to 0.76.

DISCUSSION

In this comparative assessment of breast cancer risk models, all AI algorithms had significantly higher discrimination than the BCSC clinical risk model for predicting 5-year risk. This difference was most pronounced for interval cancer risk at 0 to 1 year. Furthermore, we demonstrate that AI algorithms mostly trained for short time horizons can predict future risk of cancer up to 5 years when no cancer is detected on mammography. The combination of BCSC clinical risk and AI further improves risk prediction above AI alone, and decreases the gap in future risk performance between AI algorithms.

Mammography AI algorithms provide a new approach for improving breast cancer risk prediction beyond classical clinical variables such as age, family history, or the traditional imaging risk biomarker of breast density. The absolute increase in the AUC for the best mammography AI relative to BCSC was 0.09 for interval cancer risk and 0.06 for overall 5-year risk, which suggests that mammography AI provides 2-4 times new and independent predictors

than clinical risk factors in the BCSC model.[29] Improved interval cancer risk prediction is expected given most AI are trained for short time horizons, to aid radiologists from missing cancers. However, continued strong predictive performance up to 5 years is surprising and suggests AI is no longer identifying missed cancers, but imaging features of true underlying risk. This is analogous to breast density predicting interval cancer risk due primarily to tissue masking, but also predicting future risk where masking is no less contributory[30]. We demonstrate that AI provides prediction better than and additive with breast density, which is part of the BCSC model.

Creating a combined AI and clinical risk model demonstrated a significant, albeit slight improvement in performance compared with any AI model alone. This incremental improvement was also noted in other studies combining mammography AI and clinical risk.[12,13] The combined model also decreased overall differences in discrimination between AI algorithms. Larger gains in improvement may be derived by combining clinical risk and mammography AI with single nucleotide polymorphism polygenic risk scores,[12] which we intend to evaluate in a future study. We evaluated risk at different time horizons because each has distinct clinical implications. AI algorithms particularly excel at predicting high risk of interval cancer, which is associated with aggressive cancers[30,31] and may lead to second reading of mammogram, supplementary screening (eg, with breast MRI), or short-interval follow-up. AI algorithms also predict elevated future risk, which may lead to more frequent and intensive screening or risk counseling for primary prevention.

The BCSC model prediction was calibrated to US national SEER cancer incidence rates and remained well calibrated in our cohort, confirming that our population is likely representative of community-based populations. In contrast, the Mirai model overestimated

cancer risk by a factor of 2 across all risk strata. The Mirai predicted risk was originally calibrated using women from a tertiary referral setting who likely had a higher cancer rate than the women in our health system. Although calibration does not affect the observed discriminative performance, it is critical when clinical decisions are based on prespecified risk model thresholds. However, given its systematic overestimation, the Mirai model may be recalibrated for these purposes.

Beyond improved performance, mammography-based AI risk models provide practical advantages over traditional clinical risk models. AI uses a single data source (the screening mammogram) that is available for most women for whom breast cancer risk prediction is relevant. As a result, risk scores can be generated consistently and efficiently for all women in a large population. Mammography AI risk models overcome certain barriers for risk models such as time and cost for combining multiple data elements from potentially different sources, as well as dependence on patient-reported history, and susceptibility to missing data or recall bias. However, mammography AI risk models also entail their own challenges in terms of potential costs (eg, new software or graphics processing hardware) and other new technical and workflow considerations for implementation. Some breast imaging practices may already incorporate mammography AI trained for aiding immediate image detection, and the score that is generated can simultaneously be used for future risk stratification. However, before AI is applied, it should be evaluated in specific populations to evaluate hidden biases that may create health equity disparities in certain groups.[32]

We evaluated our results using a community-based, diverse cohort, using a rigorous design and methods to evaluate AI under both pragmatic and optimal conditions. Our observed discrimination was consistent with prior publications for the Mirai AI algorithm[13] and the BCSC

clinical risk model.[6,7,33] Our results are not an endorsement of any one AI algorithm, but a demonstration of the inherent predictive power in mammography-based deep-learning using a sample of 5 AI algorithms. It is beyond the scope of our study to evaluate the hundreds of mammography AI algorithms available at this time,[11] but similar results may be seen in algorithms not evaluated in our study.

Our study was limited due to retrospective ascertainment of BCSC clinical risk model inputs for family history and prior breast biopsies. We are unable to assess the extent to which these data are missing, particularly for breast biopsies performed prior to enrollment in our health plan. Although family history data were comparable to BCSC estimates, breast biopsy history was 10-15% lower than previously reported,[7] which may contribute to underestimation of BCSC performance. Additionally, COVID shelter-in-place orders likely decreased baseline cancer incidence in the final year of our study due to decreased screening.

Our results imply that mammography AI alone may be a powerful, step-wise improvement over clinical risk models at early time horizons, with further improved prediction when both AI and clinical risk models are combined. Although AI performance declines with longer time horizons, most of the algorithms evaluated have not been yet trained to predict longer-term outcomes, suggesting a rich opportunity for further improvement. Moreover, AI provides an especially powerful way to stratify women for clinical considerations that necessitate shorter time horizons, such as risk-based screening and supplemental imaging. The impact on clinical decisions requiring longer-term data, such as for chemoprevention or hereditary genetic screening, requires further study in cohorts with longer follow-up.

REFERENCES

1. Pfeiffer RM, Gail MH. *Absolute Risk: Methods and Applications in Clinical Management and Public Health*. 2017. Accessed September 15, 2021. https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=4943927

2. Shieh Y, Eklund M, Madlensky L, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. J Natl Cancer Inst. 2017;109(5):djw290.

3. Pashayan N, Antoniou AC, Ivanus U, et al. Personalized Early Detection and Prevention of Breast Cancer: ENVISION Consensus Statement. Nat Rev Clin Oncol. 2020;17(11):687-705.

4. Miglioretti DL, Bissell MCS, Kerlikowske K, et al. Assessment of a Risk-Based Approach for Triaging Mammography Examinations During Periods of Reduced Capacity. JAMA Netw Open. 2021;4(3):e211974.

5. Gail MH, Brinton LA, Byar DP, et al. Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. J Natl Cancer Inst. 1989;81(24):1879-86.

6. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using Clinical Factors and Mammographic Breast Density to Estimate Breast Cancer Risk: Development and Validation of a New Predictive Model. Ann Intern Med. 2008;148(5):337–47.

7. Tice JA, Miglioretti DL, Li CS, Vachon CM, Gard CC, Kerlikowske K. Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. J Clin Oncol. 2015;33(28):3137-43.

8. Tyrer J, Duffy SW, Cuzick J. A Breast Cancer Prediction Model Incorporating Familial and Personal Risk Factors. Statist Med. 2004;23(7):1111-30.

9. Rizzo S, Botta F, Raimondi S, et al. Radiomics: The Facts and the Challenges of Image Analysis. Eur Radiol Exp. 2018;2(1):36.

10. Geras KJ, Mann RM, Moy L. Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Concepts and Future Perspectives. Radiology. 2019;293(2):246-59.

11. Schaffter T, Buist DSM, Lee CI, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. JAMA Netw Open. 2020;3(3):e200265.

12. Eriksson M, Czene K, Strand F, et al. Identification of Women at High Risk of Breast Cancer Who Need Supplemental Screening. Radiology. 2020;297(2):327-33.

13. Yala A, Mikhael PG, Strand F, et al. Toward Robust Mammography-Based Models for Breast Cancer Risk. Sci Transl Med. 2021;13(578):eaba4373.

14. Daly MB, Pal T, Berry MP, et al. Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2021;19(1):77-102.

15. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. J Clin Epidemiol. 2008;61(4):344-9.

16. Sharp SJ, Poulaliou M, Thompson SG, White IR, Wood AM. A Review of Published Analyses of Case-Cohort Studies and Recommendations for Future Reporting. Gagnier JJ, ed. PLoS ONE. 2014;9(6):e101176.

17. Callahan M, Sanderson J. A Breast Cancer Tracking System. Perm J. 2000;4(4):36-9.

18. NCI-Funded Breast Cancer Surveillance Consortium (P01 CA154292 and HHSN261201100031C). 2021. Accessed August 18, 2021. https://tools.bcsc-scc.org/BC5yearRisk/sourcecode.htm

19. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2019. https://www.R-project.org

20. Heagerty PJ, Lumley T, Pepe MS. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. Biometrics. 2000;56(2):337-44.

21. Liu D, Cai T, Zheng Y. Evaluating the Predictive Value of Biomarkers with Stratified Case-Cohort Design. Biometrics. 2012;68(4):1219-27.

22. Efron B, Tibshirani R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Statist Sci. 1986;1(1):54-75.

23. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and Comparing Time-Dependent Areas under Receiver Operating Characteristic Curves for Censored Event Times with Competing Risks. Statist Med. 2013;32(30):5381-97.

24. Harrell Jr. FE. Rms: Regression Modeling Strategies. 2020. https://CRAN.R-project.org/package=rms

25. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer International Publishing; 2015.

26. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density. JAMA Oncol. 2018;4(9):e180174.

27. Liddell FD. Simple Exact Analysis of the Standardised Mortality Ratio. J Epidemiol Community Health. 1984;38(1):85-8.

28. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Wolters Kluwer Health; 2015. Accessed September 14, 2021. http://public.eblib.com/choice/publicfullrecord.aspx?p=2032120

29. Gail MH, Pfeiffer RM. Breast Cancer Risk Model Requirements for Counseling, Prevention, and Screening. J Natl Cancer Inst. 2018;110(9):994-1002.

30. Kerlikowske K, Zhu W, Tosteson ANA, et al. Identifying Women With Dense Breasts at High Risk for Interval Cancer: A Cohort Study. Ann Intern Med. 2015;162(10):673-81.

31. Porter PL, El-Bastawissi AY, Mandelson MT, et al. Breast Tumor Characteristics as Predictors of Mammographic Detection: Comparison of Interval- and Screen-Detected Cancers. J Natl Cancer Inst. 1999;91(23):2020-8.

32. Wiens J, Saria S, Sendak M, et al. Do No Harm: a Roadmap for Responsible Machine Learning for Health Care. Nat Med. 2019;25(9):1337-40.

33. Tice JA, Bissell MCS, Miglioretti DL, et al. Validation of the Breast Cancer Surveillance Consortium Model of Breast Cancer Risk. Breast Cancer Res Treat. 2019;175(2):519-23.

Table 3.1: Missing data by model

Cancer Risk Type		All observations	All missing (%)	Cases	Cases missing (%)	Non cases	Non cases missing (%)
Clinical Risk	BCSC						
AI CAD and Risk	Mirai (MIT)	18356	602 (3.3)	231	231 (4.9)	371	371 (2.7)
AI CAD	GMIC (NYU)	18356	352 (1.9)	245	245 (5.2)	107	107 (0.8)
	Vendor A	18356	479 (2.6)	153	153 (3.3)	326	326 (2.4)
	Vendor B	18356	2407 (13.1)	679	679 (14.5)	1728	1728 (12.6)
	Vendor D	18356	625 (3.4)	229	229 (4.9)	396	396 (2.9)

Abbreviations: AI: Artificial Intelligence; AUC(t): time-varying area under the curve; BCSC: Breast Cancer Surveillance Consortium; CAD: Computer-aided detection; GMIC: Globally-Aware Multiple Instance Classifier

Table 3.2: Complete cases across all models only

		Cancer Risk Type	Interval cancer risk	Future cancer risk (excluding interval cancers at 0 to 1 year)				All cancer risk
	Time Horizon / Evaluated		0 to 1 year	1 to 2 years	1 to 3 years	1 to 4 years	1 to 5 years	0 to 5 years
	Model / Trained							
Clinical Risk	BCSC	0 to 5 years	0.63 (0.58,0.66)	0.62 (0.6,0.63)	0.63 (0.61,0.64)	0.62 (0.6,0.63)	0.61 (0.6,0.62)	0.61 (0.6,0.62)
AI CAD and Risk	Mirai (MIT)	0 to 5 years	0.69 (0.64,0.71)	0.69 (0.66,0.7)	0.69 (0.68,0.7)	0.69 (0.67,0.69)	0.68 (0.66,0.68)	0.68 (0.66,0.68)
AI CAD	GMIC (NYU)	0 to 0.25 years	0.68 (0.64,0.71)	0.67 (0.64,0.68)	0.67 (0.66,0.68)	0.66 (0.65,0.67)	0.64 (0.63,0.65)	0.65 (0.63,0.65)
	Vendor A	0 to 2 years	0.72 (0.68,0.75)	0.69 (0.67,0.71)	0.69 (0.67,0.69)	0.67 (0.66,0.68)	0.65 (0.64,0.66)	0.66 (0.64,0.66)
	Vendor B	0 to 1 year	0.73 (0.67,0.74)	0.68 (0.64,0.68)	0.67 (0.64,0.67)	0.66 (0.63,0.65)	0.64 (0.62,0.64)	0.64 (0.62,0.64)
	Vendor D	0 to 2 years	0.67 (0.63,0.7)	0.68 (0.65,0.69)	0.68 (0.67,0.69)	0.67 (0.65,0.67)	0.65 (0.64,0.66)	0.65 (0.64,0.66)

Abbreviations: AI: Artificial Intelligence; AUC(t): time-varying area under the curve; BCSC: Breast Cancer Surveillance Consortium; CAD: Computer-aided detection; GMIC: Globally-Aware Multiple Instance Classifier

Table 3.3: Patient characteristics

	Women in subcohort, n (%)	Women with breast cancer, n (%)	All eligible women, n (%)
Total women	13 881 (100)	4672 (100)	329 814 (100)
Age, years			
<40	84 (1)	19 (<1)	2011 (1)
40-49	3149 (23)	735 (16)	74 887 (21)
50-59	4792 (35)	1331 (28)	114 780 (33)
60-69	4215 (30)	1809 (39)	99 341 (32)
≥70	1641 (12)	778 (17)	38 795 (13)
Race/Ethnicity			
Black, non-Hispanic	976 (7)	329 (7)	1257 (7)
Asian or Pacific Islander	2672 (19)	897 (19)	3426 (20)
Hispanic	2401 (17)	572 (12)	2921 (16)
Multiracial	504 (4)	163 (3)	657 (4)
Native American	53 (<1)	18 (<1)	69 (<1)
White, non-Hispanic	7037 (51)	2676 (57)	9383 (52)
Missing	238 (2)	17 (<1)	268 (1)
First-degree family history			
0	12 150 (88)	3750 (80)	275 535 (85)
1	1639 (12)	866 (19)	44 901 (14)
≥ 2	92 (1)	56 (1)	2854 (1)
Previous benign breast biopsies			
0	13 119 (95)	4160 (89)	16 774 (93)
≥1	762 (5)	512 (11)	1198 (7)
BI-RADS breast density			
Almost entirely fat	1405 (10)	250 (5)	1552 (9)
Scattered fibroglandular densities	6387 (46)	2014 (43)	8143 (46)
Heterogeneously dense	5341 (38)	2117 (45)	7194 (40)
Extremely dense	748 (5)	251 (5)	997 (6)
Missing	65 (<1)	40 (1)	95 (<1)
Cancer type			
Invasive	152 (77)	3850 (82)	3850 (82)
DCIS	45 (23)	822 (18)	822 (18)
Median follow-up interval, years (interquartile range)	5.0 (4.7 to 5.3)	2.8 (2.0 to 4.1)	5.0 (4.7 to 5.3)
Median healthcare enrollment prior to index date, years (interquartile range)	17.9 (9.7 to 19.4)	18.9 (10.7 to 19.5)	17.6 (9.2 to 19.4)

Table 3.4: Comparative AUC(t) performance of AI models and BCSC clinical risk model using 2016 screening mammogram without evidence of cancer on final imaging assessment for invasive cancer and DCIS.

		Interval cancer risk	**Future cancer risk (excluding interval cancers at 0 to 1 year)**				**All cancer risk**
Time Horizon	**Evaluated / Trained**	0 to 1 year (No. cancers = 259)	1 to 2 years (No. cancers = 883)	1 to 3 years (No. cancers = 2,224)	1 to 4 years (No. cancers = 3,087)	1 to 5 years (No. cancers = 4,168)	0 to 5 years (No. cancers = 4,430)
BCSC (Clinical)	0.5 to 5 years	0.62 (0.58,0.66)	0.62 (0.60,0.63)	0.62 (0.61,0.64)	0.61 (0.6,0.63)	0.61 (0.59,0.62)	0.61 (0.6,0.62)
Mirai (AI)	0 to 5 years	**0.68 (0.64,0.71)**	**0.68 (0.66,0.7)**	**0.69 (0.68,0.7)**	**0.68 (0.67,0.69)**	**0.67 (0.66,0.68)**	**0.67 (0.66,0.68)**
Vendor A (AI)	0 to 2 years	**0.71 (0.68,0.74)**	**0.69 (0.67,0.71)**	**0.68 (0.67,0.69)**	**0.67 (0.66,0.68)**	**0.65 (0.64,0.66)**	**0.65 (0.64,0.66)**
Vendor D (AI)	0 to 2 years	**0.66 (0.62,0.70)**	**0.67 (0.65,0.69)**	**0.68 (0.67,0.69)**	**0.66 (0.65,0.67)**	**0.65 (0.64,0.66)**	**0.65 (0.64,0.66)**
Vendor B (AI)	0 to 1 year	**0.71 (0.67,0.73)**	**0.66 (0.64,0.68)**	**0.65 (0.64,0.67)**	**0.64 (0.63,0.65)**	**0.63 (0.62,0.64)**	**0.63 (0.62,0.64)**
GMIC (AI)	0 to 0.25 years	**0.68 (0.64,0.70)**	**0.66 (0.65,0.68)**	**0.67 (0.66,0.68)**	**0.66 (0.65,0.67)**	**0.64 (0.63,0.65)**	**0.64 (0.63,0.65)**

Abbreviations: AI: Artificial Intelligence; AUC(t): time-varying area under the curve; BCSC: Breast Cancer Surveillance Consortium; CAD: Computer-aided detection; GMIC: Globally-Aware Multiple Instance Classifier

bold: ***P < .05, as compared with the BCSC clinical risk model for the same time horizon***

Table 3.5: Comparative time-varying AUC(t) performance of combined AI models and BCSC clinical risk model using 2016 screening mammogram without evidence of cancer on final imaging assessment.

		Interval cancer risk	**Future cancer risk (excluding interval cancers in 0 to 1 year)**				**All cancer risk**
Time Horizon / **Evaluated** / **Model** / **Trained**		0 to 1 year (No. cancers = 259)	1 to 2 years (No. cancers = 883)	1 to 3 years (No. cancers = 2,224)	1 to 4 years (No. cancers = 3,087)	1 to 5 years (No. cancers = 4,168)	0 to 5 years (No. cancers = 4,430)
BCSC + Mirai (MIT)	0.5 to 5 years	**0.69 (0.65,0.72)**	**0.7 (0.68,0.71)**	**0.7 (0.69,0.71)**	**0.69 (0.68,0.7)**	**0.68 (0.67,0.69)**	**0.68 (0.67,0.69)**
BCSC + Vendor A	0 to 2 years	0.73 (0.69,0.76)	**0.71 (0.69,0.73)**	**0.7 (0.69,0.72)**	**0.69 (0.68,0.7)**	**0.67 (0.66,0.68)**	**0.67 (0.66,0.68)**
BCSC + Vendor D	0 to 2 years	0.67 (0.63,0.71)	**0.68 (0.66,0.7)**	**0.68 (0.67,0.7)**	**0.67 (0.66,0.68)**	**0.65 (0.64,0.66)**	**0.66 (0.65,0.67)**
BCSC + Vendor B	0 to 1 year	0.72 (0.68,0.75)	**0.7 (0.68,0.71)**	**0.69 (0.68,0.7)**	**0.67 (0.66,0.68)**	**0.66 (0.65,0.67)**	**0.66 (0.65,0.67)**
BCSC + GMIC (NYU)	0 to 0.25 years	**0.69 (0.65,0.72)**	**0.69 (0.67,0.71)**	**0.69 (0.68,0.7)**	**0.68 (0.67,0.69)**	**0.66 (0.65,0.67)**	**0.66 (0.66,0.67)**

Combined model was fit using restricted cubic splines

bold: ***P* < .05, as compared with the corresponding AI only model (Table 2) for the same time horizon**

Table 3.6: Invasive cancer only. Comparative AUC(t) performance of AI models and BCSC clinical risk model for invasive cancer only using 2016 screening mammogram without evidence of cancer on final imaging assessment

		Cancer Risk Type*	**Interval cancer risk**	**Future cancer risk (excluding interval cancers at 0 to 1 year)**				**All cancer risk**
	Time Horizon / **Evaluated** / **Model** / **Trained**		0 to 1 year	1 to 2 years	1 to 3 years	1 to 4 years	1 to 5 years	0 to 5 years
Clinical Risk	BCSC	0 to 5 years	0.61	0.61	0.63	0.62	0.61	0.61
AI CAD and Risk	Mirai (MIT)	0 to 5 years	0.67	0.67	0.68	0.68	0.67	0.67
AI CAD	GMIC (NYU)	0 to 2.5 years	0.66	0.64	0.66	0.65	0.63	0.64
	Vendor A	0 to 2 years	0.70	0.68	0.68	0.67	0.65	0.65
	Vendor B	0 to 1 year	0.70	0.66	0.66	0.64	0.63	0.63
	Vendor D	0 to 2 years	0.65	0.66	0.68	0.66	0.65	0.65

Abbreviations: AI: Artificial Intelligence; BCSC: Breast Cancer Surveillance Consortium; CAD: Computer-aided detection; GMIC: Globally Aware Multiple Instance Classifier

Table 3.7: Screening mammograms acquired on Hologic equipment, complete cases only

		Cancer Risk Type	Interval cancer risk	Future cancer risk (excluding interval cancers at 0 to 1 year)				All cancer risk
	Time Horizon / Evaluated	Trained	0 to 1 year	1 to 2 years	1 to 3 years	1 to 4 years	1 to 5 years	0 to 5 years
	Model							
Clinical Risk	BCSC	0 to 5 years	0.63 (0.59,0.67)	0.63 (0.6,0.65)	0.63 (0.62,0.65)	0.62 (0.61,0.63)	0.61 (0.6,0.62)	0.61 (0.6,0.62)
AI CAD and Risk	Mirai (MIT)	0 to 5 years	0.69 (0.65,0.72)	0.69 (0.67,0.71)	0.69 (0.68,0.71)	0.69 (0.68,0.7)	0.67 (0.66,0.69)	0.68 (0.66,0.69)
AI CAD	GMIC (NYU)	0 to 0.25 years	0.69 (0.65,0.72)	0.67 (0.65,0.69)	0.68 (0.66,0.69)	0.67 (0.66,0.68)	0.65 (0.64,0.66)	0.66 (0.64,0.67)
	Vendor A	0 to 2 years	0.72 (0.68,0.76)	0.69 (0.67,0.72)	0.68 (0.67,0.7)	0.67 (0.66,0.68)	0.65 (0.64,0.66)	0.66 (0.65,0.67)
	Vendor B	0 to 1 year	0.73 (0.69,0.77)	0.68 (0.66,0.7)	0.67 (0.66,0.69)	0.66 (0.65,0.67)	0.65 (0.63,0.66)	0.65 (0.64,0.66)
	Vendor D	0 to 2 years	0.67 (0.63,0.71)	0.68 (0.65,0.7)	0.68 (0.67,0.69)	0.66 (0.65,0.68)	0.65 (0.64,0.66)	0.65 (0.64,0.66)

Abbreviations: AI: Artificial Intelligence; AUC(t): time-varying area under the curve; BCSC: Breast Cancer Surveillance Consortium; CAD: Computer-aided detection; GMIC: Globally-Aware Multiple Instance Classifier

Table 3.8: Screening mammograms acquired on General Electric equipment, complete cases only

		Cancer Risk Type	Interval cancer risk	Future cancer risk (excluding interval cancers at 0 to 1 year)				All cancer risk
	Time Horizon / Evaluated / Model / Trained		0 to 1 year	1 to 2 years	1 to 3 years	1 to 4 years	1 to 5 years	0 to 5 years
Clinical Risk	BCSC	0 to 5 years	0.60 (0.46,0.71)	0.61 (0.56,0.67)	0.62 (0.58,0.66)	0.61 (0.58,0.64)	0.61 (0.58,0.64)	0.61 (0.58,0.64)
AI CAD and Risk	Mirai (MIT)	0 to 5 years	0.69 (0.56,0.79)	0.66 (0.59,0.72)	0.7 (0.66,0.73)	0.7 (0.66,0.72)	0.67 (0.64,0.7)	0.67 (0.65,0.7)
AI CAD	GMIC (NYU)	0 to 0.25 years	0.63 (0.50,0.73)	0.61 (0.55,0.67)	0.65 (0.61,0.68)	0.63 (0.60,0.67)	0.61 (0.58,0.64)	0.61 (0.58,0.64)
	Vendor A	0 to 2 years	0.72 (0.60,0.81)	0.68 (0.61,0.74)	0.69 (0.66,0.73)	0.68 (0.65,0.71)	0.65 (0.61,0.68)	0.65 (0.62,0.68)
	Vendor B	0 to 1 year	0.74 (0.56,0.85)	0.67 (0.58,0.74)	0.68 (0.62,0.72)	0.66 (0.6,0.7)	0.62 (0.58,0.67)	0.63 (0.59,0.68)
	Vendor D	0 to 2 years	0.70 (0.56,0.79)	0.69 (0.62,0.75)	0.72 (0.68,0.75)	0.69 (0.65,0.72)	0.67 (0.64,0.7)	0.67 (0.64,0.7)

Abbreviations: AI: Artificial Intelligence; AUC(t): time-varying area under the curve; BCSC: Breast Cancer Surveillance Consortium; CAD: Computer-aided detection; GMIC: Globally-Aware Multiple Instance Classifier

Table 3.9: Screening BI-RADS 1 or 2 only

		Cancer Risk Type	Interval cancer risk	Future cancer risk (excluding interval cancers at 0 to 1 year)				All cancer risk
	Time Horizon / Evaluated / Model / Trained		0 to 1 year	1 to 2 years	1 to 3 years	1 to 4 years	1 to 5 years	0 to 5 years
Clinical Risk	BCSC	0 to 5 years	0.63 (0.59,0.66)	0.62 (0.6,0.64)	0.63 (0.61,0.64)	0.62 (0.61,0.63)	0.61 (0.6,0.62)	0.61 (0.6,0.62)
AI CAD and Risk	Mirai (MIT)	0 to 5 years	0.68 (0.64,0.71)	0.69 (0.67,0.71)	0.69 (0.68,0.7)	0.69 (0.68,0.7)	0.67 (0.66,0.68)	0.67 (0.66,0.68)
AI CAD	GMIC (NYU)	0 to 0.25 years	0.67 (0.64,0.71)	0.67 (0.65,0.69)	0.67 (0.66,0.68)	0.66 (0.65,0.67)	0.64 (0.63,0.66)	0.65 (0.64,0.66)
	Vendor A	0 to 2 years	0.71 (0.67,0.74)	0.69 (0.67,0.71)	0.68 (0.67,0.69)	0.67 (0.66,0.68)	0.65 (0.64,0.66)	0.65 (0.64,0.66)
	Vendor B	0 to 1 year	0.70 (0.67,0.73)	0.66 (0.64,0.68)	0.65 (0.64,0.67)	0.64 (0.63,0.65)	0.63 (0.62,0.64)	0.63 (0.62,0.64)
	Vendor D	0 to 2 years	0.66 (0.62,0.7)	0.67 (0.65,0.69)	0.68 (0.67,0.69)	0.66 (0.65,0.67)	0.65 (0.64,0.66)	0.65 (0.64,0.66)

Abbreviations: AI: Artificial Intelligence; AUC(t): time-varying area under the curve; BCSC: Breast Cancer Surveillance Consortium; CAD: Computer-aided detection; GMIC: Globally-Aware Multiple Instance Classifier

Table 3.10: Screening BI-RADS 0, and Diagnostic or post-procedural BI-RADS 1 or 2 only

		Cancer Risk Type	**Interval cancer risk**	**Future cancer risk (excluding interval cancers at 0 to 1 year)**				**All cancer risk**
	Time Horizon / **Evaluated** / **Model** / **Trained**		0 to 1 year	1 to 2 years	1 to 3 years	1 to 4 years	1 to 5 years	0 to 5 years
Clinical Risk	BCSC	0 to 5 years	0.58 (0.41,0.71)	0.62 (0.55,0.68)	0.63 (0.58,0.68)	0.62 (0.57,0.66)	0.60 (0.56,0.65)	0.60 (0.56,0.65)
AI CAD and Risk	Mirai (MIT)	0 to 5 years	0.67 (0.52,0.79)	0.61 (0.54,0.67)	0.64 (0.58,0.68)	0.65 (0.6,0.69)	0.66 (0.61,0.69)	0.66 (0.62,0.7)
AI CAD	GMIC (NYU)	0 to 0.25 years	0.64 (0.48,0.75)	0.56 (0.48,0.62)	0.62 (0.57,0.67)	0.53 (0.49,0.57)	0.54 (0.49,0.58)	0.54 (0.5,0.59)
	Vendor A	0 to 2 years	0.75 (0.61,0.84)	0.63 (0.56,0.69)	0.68 (0.63,0.72)	0.66 (0.62,0.7)	0.65 (0.6,0.69)	0.66 (0.61,0.7)
	Vendor B	0 to 1 year	0.74 (0.57,0.83)	0.67 (0.6,0.73)	0.63 (0.58,0.68)	0.6 (0.56,0.64)	0.61 (0.57,0.65)	0.62 (0.58,0.66)
	Vendor D	0 to 2 years	0.66 (0.51,0.76)	0.63 (0.55,0.69)	0.66 (0.6,0.71)	0.65 (0.6,0.69)	0.65 (0.61,0.69)	0.65 (0.61,0.69)

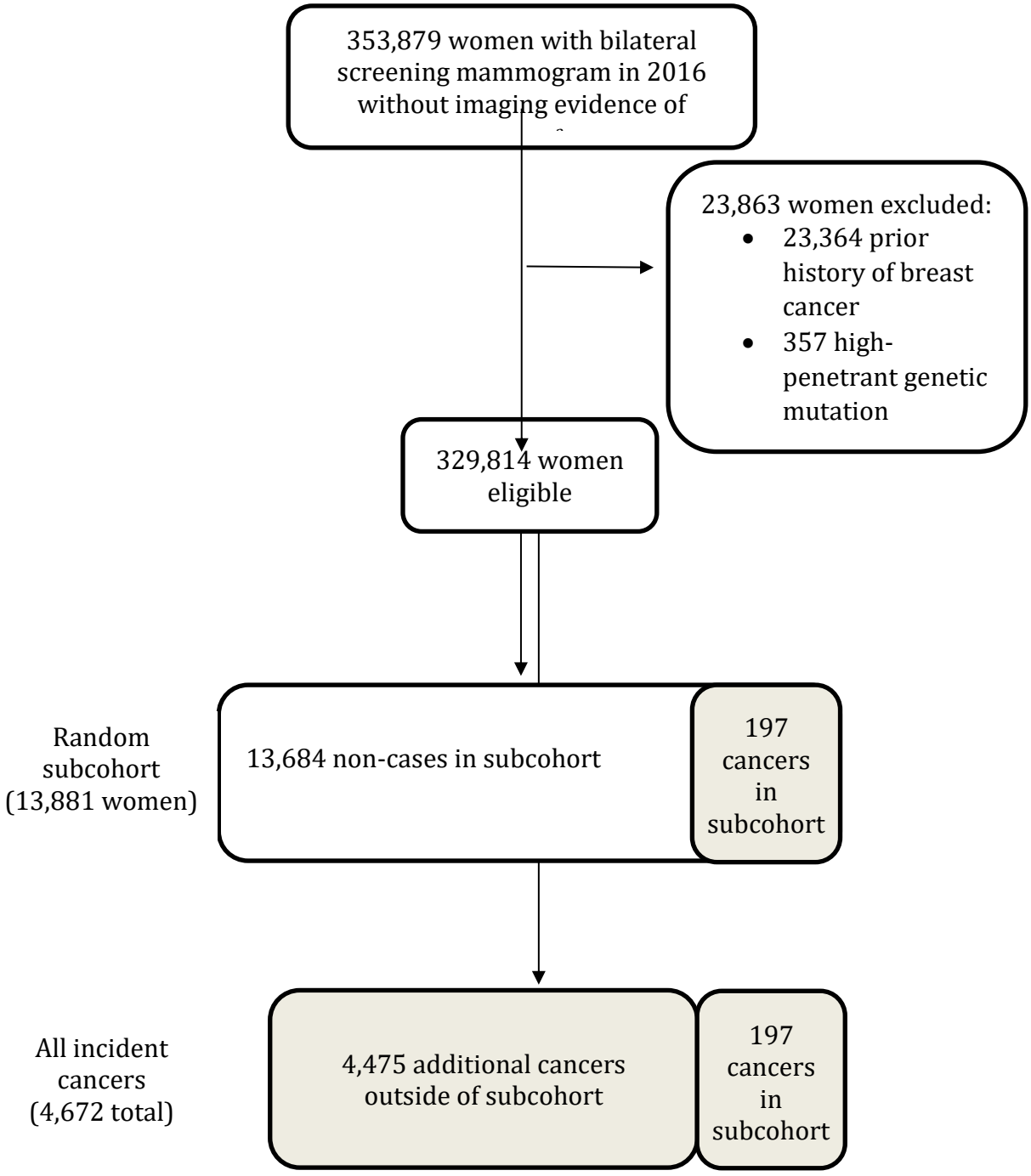
Abbreviations: AI: Artificial Intelligence; AUC(t): time-varying area under the curve; BCSC: Breast Cancer Surveillance Consortium; CAD: Computer-aided detection; GMIC: Globally-Aware Multiple Instance Classifier

Table 3.11: Absolute risk calibration by Mirai AI and BCSC clinical risk models at 5-year cumulative incidence thresholds

			No. of Cancer Cases			IR per 1000 women/y		
Model by 5-year Risk	No. (%) of women	Follow-up, 1000 women-years (weighted)	Observed	Expected (weighted)	O/E (95% CI)	Observed	Expected (weighted)	IRR (95% CI)
BCSC								
All	18 356	1422.7	4430	4172.4	1.06 (1.03,1.09)	3.11	2.93	
0-<1.67%	14 499	1164.5	3048	2825.2	1.08 (1.04,1.12)	2.62	2.43	1.00
1.67-<3%	3387	233.3	1150	1129.5	1.02 (0.96,1.08)	4.93	4.84	1.88 (1.76,2.02)
≥3%	470	25.0	232	217.7	1.07 (0.93,1.21)	9.28	8.71	3.55 (3.10,4.05)
Mirai								
All	18 356	1422.7	4430	7360.5	0.60 (0.58,0.62)	3.11	5.17	
0-<1.67%	7597	672.4	1010	2060.1	0.49 (0.46,0.52)	1.50	3.06	1.00
1.67-<3%	8042	592.6	2233	2935.9	0.76 (0.73,0.79)	3.77	4.95	2.51 (2.33,2.70)
≥3%	2717	157.6	1187	2364.5	0.5 (0.47,0.53)	7.53	15.00	5.01 (4.61,5.45)

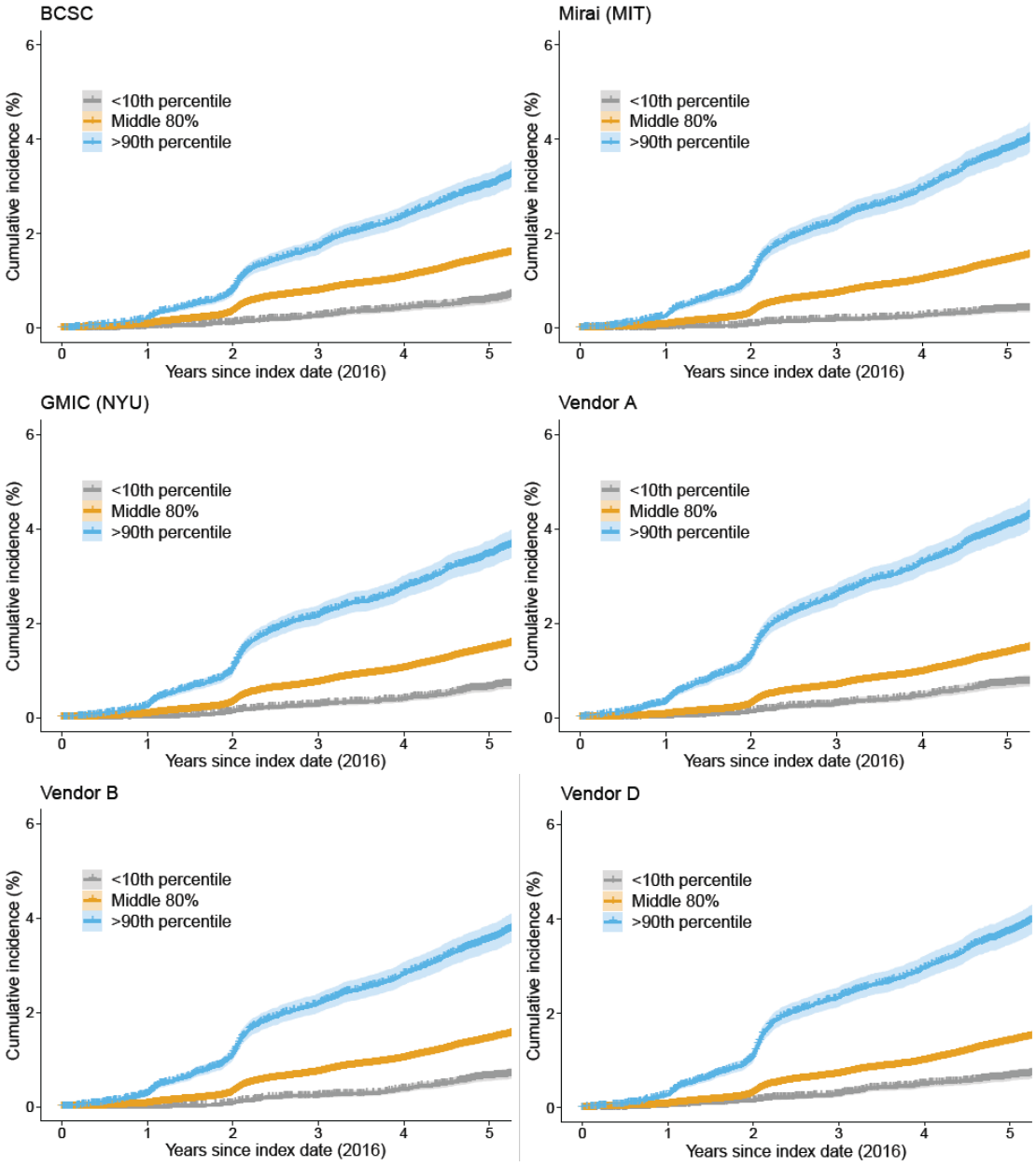
Abbreviations: O/E: Observed to expected ratio; IR: Incidence rate; IRR: Incidence rate ratio; BCSC: Breast Cancer Surveillance Consortium

Figure 3.1: Case cohort selection



[a] No imaging evidence of cancer: Screening examination BI-RADS 1 or 2 OR Screening BI-RADS 0 and diagnostic BI-RADS 1 or 2 in ≤ 90 days OR Screening BI-RADS 0 and diagnostic BI-RADS 4 or 5 and benign biopsy in ≤ 90 days

Figure 3.2: Cumulative risk of breast cancer by AI score score at 5 years



Appendix 3.1: Description of Artificial Algorithms

Survey questions were submitted to algorithm developers to further describe the underlying training, data, and architecture.

Globally-Aware Multiple Instance Classifier (GMIC, NYU):

1. **Briefly describe the demographic and platform-specific composition of the data used for training: Total training sample size? Does the data contain non-white patients? Are the mammograms from multiple imaging platforms?**

We used NYU Breast Cancer Screening Dataset to develop our model. This dataset contains 229,426 exams (1,001,093 images) from 141,472 patients who imaged at NYU Langone Health between 2010 and 2017. The dataset contains non-white patients. You can find more information in this tech report.

2. **Briefly describe the types or combination of the types of deep learning algorithms used by your model: e.g., CNN, DNN, GAN, etc.? Does your model utilize pretrained models or transfer learning (e.g., resnet, inception, etc.)?**

The primary methodologies applied in this paper: CNN, weakly supervised learning. We used ResNet pretraining weights from ImageNet.

3. Briefly describe the core technologies and/or framework(s) used by your model (e.g., Python, Java/JVM, TensorFlow, Pytorch, Caffe, etc.)

PyTorch

4. Can your model process mammography studies beyond standard 4 views (e.g., unilateral only, more than 4 views, implant)?

Yes. It doesn't make any assumption on the view.

5. How were the positive and negative labels of training images defined (e.g., specify the time interval from image to diagnosis, pathologically confirmed invasive cancer or DCIS, benign lesions included among negatives or as a third outcome, no known breast cancer diagnosis within 5 years, one or two subsequent negative screening exams)?

A breast was defined as cancer-positive if there was at least one pathology report confirming the presence of malignant lesion within 120 days of the time when the mammography images were acquired. DCIS was considered as malignant. Benign findings such as cyst and fibroadenoma were treated as another class. Cancer negative exams include exams with benign/normal findings or exams that weren't escalated for biopsy. Please find more information in the tech report (referenced below) of this dataset.

6. Briefly describe your model's input requirement (e.g., DICOM, PNG, "for presentation" or "for processing" view). Is any preprocessing required?

The implementation published on GitHub takes in a 2D / 3D matrix representation of a mammography image. The user needs to extract the image data from DICOM. Image pre-processing is included as part of the git repo.

7. Briefly describe your model's output (e.g., does it represent the probability of cancer or can it be converted to a probability?). Does the output include anything else, such as bounding boxes or lesion segmentation?

The model outputs two probability scores on the presence of any benign and malignant lesion in a mammography image. The model also returns saliency maps which highlight the areas on the images that could correspond to a benign/malignant lesion. See Figure 7 of our paper as an example.

8. How does your model generate breast-level or patient-level predictions?

Breast-level predictions were calculated as the simple average over all image-level predictions.

9. Does your model employ any inference-time techniques, such as model ensemble or data augmentation?

We use model ensembling and test time data augmentation. See more details in Section 3.3.1 of our paper.

10. Is your model able to consider any prior exams?
No

11. Relevant citations?

1. The paper: https://www.sciencedirect.com/science/article/pii/S1361841520302723
2. NYU Breast Cancer Screening Dataset: https://cs.nyu.edu/~kgeras/reports/datav1.0.pdf
3. The original version of our model: https://link.springer.com/chapter/10.1007/978-3-030-32692-0_3

Mirai (MIT):

1. **Briefly describe the demographic and platform-specific composition of the data used for training: Total training sample size? Does the data contain non-white patients? Are the mammograms from multiple imaging platforms?**

The full demographics of our training set are in table 3 of the paper.

2. Briefly describe the types or combination of the types of deep learning algorithms used by your model: e.g., CNN, DNN, GAN, etc.? Does your model utilize pretrained models or transfer learning (e.g., resnet, inception, etc.)?

This is detailed in our paper.
Mirai leverages a ResNet to encode individual views, and a Transformer to combine multiple view representations into a patient level representation. The model was trained to predict multiple time-points simultaneously using our Additive Hazard layer, to predict traditional clinical risk factors (e.g. age) from the image. To make our model consistent across different mammography machines in our dataset, we used conditional adversarial training.

3. Briefly describe the core technologies and/or framework(s) used by your model (e.g., Python, Java/JVM, TensorFlow, Pytorch, Caffe, etc.)
Mirai was built in PyTorch and Python

4. Can your model process mammography studies beyond standard 4 views (e.g., unilateral only, more than 4 views, implant)?
Our model requires all four standard views (R CC, R MLO, L CC, L MLO)

5. How were the positive and negative labels of training images defined (e.g., specify the time interval from image to diagnosis, pathologically confirmed invasive cancer or DCIS, benign lesions included among negatives or as a third outcome, no known breast cancer diagnosis within 5 years, one or two subsequent negative screening exams)?
We trained our model to predict cancer across multiple timepoints. A patient was considered "positive" for cancer within three years if they had a pathologically confirmed invasive cancer or DCIS diagnosis within three years of their mammogram. A patient was considered "negative" for cancer within three years if they had at least three years of screening followup without such a diagnosis. We didn't exclude benign lesions.

6. Briefly describe your model's input requirement (e.g., DICOM, PNG, "for presentation" or "for processing" view). Is any preprocessing required?
Our code assumes For Presentation dicoms, and will convert them to pngs using the DCMTK library

7. Briefly describe your model's output (e.g., does it represent the probability of cancer or can it be converted to a probability?). Does the output include anything else, such as bounding boxes or lesion segmentation?
The model outputs the probability of a cancer diagnosis within one to five years. This is represented as a 5 dimensional probability array.

8. How does your model generate breast-level or patient-level predictions?
The model generates patient-level predictions.

9. Does your model employ any inference-time techniques, such as model ensemble or data augmentation?
We do not apply model ensembling or test-time data augmentation.

10. Is your model able to consider any prior exams?
The model does not leverage prior mammograms in its predictions.

11. Relevant citations?

Yala A, Mikhael PG, Strand F, et al. Toward robust mammography-based models for breast cancer risk. *Sci Transl Med*. 2021;13(578):eaba4373. doi:10.1126/scitranslmed.aba4373

Vendor A:

1. **Briefly describe the demographic and platform-specific composition of the data used for training: Total training sample size? Does the data contain non-white patients? Are the mammograms from multiple imaging platforms?**
We collected and curated a dataset of 1.3M images originating from Europe (France, UK) and the USA. It covers a wide range of imaging platforms, the most prominent being: Hologic, GE, Fuji, Giotto, Siemens, Philips. Images are a mixture of FFDM, DBT and 2DSM (synthetic mammography). We know from the collected centers that non-white patients are included without being able to determine precisely how many.

2. **Briefly describe the types or combination of the types of deep learning algorithms used by your model: e.g., CNN, DNN, GAN, etc.? Does your model utilize pretrained models or transfer learning (e.g., resnet, inception, etc.)?**
We use a mixture of 5 families of convolutional neural networks (CNN), each having a specific purpose. A first type of CNN takes a whole mammographic view as input and outputs its likelihood of malignancy. A second type of CNN extends the first one by leveraging the (lack of) symmetry between a view and its symmetrical counterpart. A third CNN is specialized in detecting all anomalies in a view, regardless of their likelihood of malignancy. A fourth CNN, further extended by a final CNN leveraging the (lack of) symmetry, takes as input high-resolution patches around detections obtained by the previous CNN and characterizes their level of suspicion. The final output of the algorithm consists in a set of positions (coordinates) within each breast view with their consolidated likelihood of malignancy obtained by fusing the image-wise and patch-wise predictions. Each model family comes with 10 instances trained by cross-validation, making a total of 50 CNNs executed on each view of each mammogram. Findings in cranial and lateral views of the

same laterality are finally paired using iconic and geometrical heuristics for a more consistent output.

3. Briefly describe the core technologies and/or framework(s) used by your model (e.g., Python, Java/JVM, TensorFlow, Pytorch, Caffe, etc.)

The primary programming language is Python. The deep learning framework used is TensorFlow version 2. Additional machine learning methods from scikit-learn and XGBoost are also used.

4. Can your model process mammography studies beyond standard 4 views (e.g., unilateral only, more than 4 views, implant)?

We currently support a maximum of 4 views per mammogram, but we do support less (though off-label). Unilateral mammograms (e.g., L-CC and L-MLO) are supported (symmetric models are disabled in this case). Unique views (e.g., L-CC and R-CC) are also supported. Additional screening views such as ML, LM, XCC are also supported. In case of duplicated views (e.g., two L-CC), the most recent image is selected (we hypothesize that in case of duplicated views the most recent is a reshoot due to bad quality of the older version).

5. **How were the positive and negative labels of training images defined (e.g., specify the time interval from image to diagnosis, pathologically confirmed invasive cancer or DCIS, benign lesions included among negatives or as a third outcome, no known breast cancer diagnosis within 5 years, one or two subsequent negative screening exams)?**

Positive cases were confirmed by a positive biopsy (either invasive cancer of DCIS) within 24 months from screening date. For each mammogram, the cancer presence was confirmed (annotated) by an expert radiologist to avoid injecting interval cancers in the training set. Regarding negative cases, those were confirmed by a negative or benign screening exam within 24 months after the screening date. Cases with benign lesions (as confirmed after a diagnostic mammogram or a biopsy) were included in the same group as negative cases. Cases with known history of breast cancer (regardless of when it happened) and breast surgery were excluded.

6. **Briefly describe your model's input requirement (e.g., DICOM, PNG, "for presentation" or "for processing" view). Is any preprocessing required?**

We use the FOR PRESENTATION DICOM images for the AI analysis. No additional preprocessing is required. Images are normalized internally (i.e., by the algorithm) to make them look similar across vendors. The normalization procedure is not disclosed.

7. **Briefly describe your model's output (e.g., does it represent the probability of cancer or can it be converted to a probability?). Does the output include anything else, such as bounding boxes or lesion segmentation?**

[Vendor A algorithm] outputs 2 scores: a raw score, ranging between 0 and 1, and a discretized score on a 1-10 scale to ease its interpretation. The score was calibrated on a screening distribution with a cancer prevalence of 5:1000 so it can directly be interpreted as a probability. Additionally, [Vendor A algorithm] returns the detection bounding boxes. No pixelic lesion segmentation is performed.

8. How does your model generate breast-level or patient-level predictions?

The breast-level score is obtained as the highest score of the detected lesions in the breast. The patient-level score is the highest score of the left and right breast. Therefore, it is always possible to connect the score at any level to a specific lesion in a specific view, which eases interpretability.

9. Does your model employ any inference-time techniques, such as model ensemble or data augmentation?

Given the multiplicity of our CNN, we extensively use ensembling techniques (bagging). Each model family has 10 instances, which are combined (averaged) to form a unique prediction for that family. No test-time augmentation is done.

10. Is your model able to consider any prior exams?

The use of prior examinations is currently under development. We have very promising preliminary results on certain model families, and are currently extending it to other families.

–

Vendor B:

1. Briefly describe the demographic and platform-specific composition of the data used for training: Total training sample size? Does the data contain non-white patients? Are the mammograms from multiple imaging platforms?

In construction of the model we utilized data from four primary vendors: Hologic, GE, Giotto, and Siemens. The model has been trained on >4M images. The dataset contains Caucasian, Asian, Hispanic, and African American demographics.

2. Briefly describe the types or combination of the types of deep learning algorithms used by your model: e.g., CNN, DNN, GAN, etc.? Does your model utilize pretrained models or transfer learning (e.g., resnet, inception, etc.)?

Our models include convolutional neural networks comprising several customized architectures. We also construct an ensemble of several of these models with various training parameters to construct the final model. We use transfer learning on internal data to make our optimization process more efficient. We don't use any publicly pretrained models to avoid potential bias.

3. Briefly describe the core technologies and/or framework(s) used by your model (e.g., Python, Java/JVM, TensorFlow, Pytorch, Caffe, etc.)

The product is built as a JVM service-based tool using dcmtk to directly interface to DICOM compatible devices. The model is built using TensorFlow as the primary framework for deep learning, with several proprietary customizations to improve the performance of our models.

4. Can your model process mammography studies beyond standard 4 views (e.g., unilateral only, more than 4 views, implant)?
Mia assesses 4 standard view images [CC-L, CC-R, MLO-L, MLO-R]. When the 4 standard views are a subset of more than 4 images in a case, Mia utilizes a custom image selection heuristic to determine an overall result based on the appropriate 4-image subset. The heuristic can be customized to fit local processes. We believe it is critical that our deployments are executed with understanding of how each local site works (e.g. how the images are produced). This information is then used to ensure Mia is optimally deployed. Versions of this heuristic have already been used successfully in independent analyses.

5. How were the positive and negative labels of training images defined (e.g., specify the time interval from image to diagnosis, pathologically confirmed invasive cancer or DCIS, benign lesions included among negatives or as a third outcome, no known breast cancer diagnosis within 5 years, one or two subsequent negative screening exams)?
Our positive definition is based on pathology or surgical follow up within 6 months or 12 months as proof of malignancy. Our negative definition has a requirement of negative at screening plus 24 to 36 months follow up with a negative result.

6. Briefly describe your model's input requirement (e.g., DICOM, PNG, "for presentation" or "for processing" view). Is any preprocessing required?
We only report on diagnosis-grade "For Presentation" Images as described in the DICOM standard. No preprocessing is required.

7. Briefly describe your model's output (e.g., does it represent the probability of cancer or can it be converted to a probability?). Does the output include anything else, such as bounding boxes or lesion segmentation?
We provide a binary decision of Recall or No Recall based on a calibration step with each provider's local data environment. Included in a recall decision would be side-wise (L/R) and view-wise (MLO/CC) details. We also provide an explanatory ROI for informing the logic behind the recall decision. We can provide a score, however it is our belief that the score is not reflective of the probability of cancer.

8. How does your model generate breast-level or patient-level predictions?
We generate a prediction at the image level across an ensemble of models and then apply a reduction heuristic to achieve the final result. The recall decision includes a case-wise, side-wise and view-wise output.

9. Does your model employ any inference-time techniques, such as model ensemble or data augmentation?

We use a varied ensemble of ML models. We found that, with our models, inference-time techniques such as test time augmentation were not needed to achieve very good performance. This allows us to be time and cost efficient at inference time.

10. Is your model able to consider any prior exams?
Some of our internal models have been trained using priors and we have seen strong performance improvements. The production model doesn't not consider any prior exams at

inference-time yet but this feature is actively being worked on and is expected to be released in 2022.

Vendor D:

1. **Briefly describe the demographic and platform-specific composition of the data used for training: Total training sample size? Does the data contain non-white patients? Are the mammograms from multiple imaging platforms?**

Predominantly Caucasian women were included in the training. Approximately one thousand breast cancer cases and ten thousand controls were used from multiple mammography machine vendors.

2. **Briefly describe the types or combination of the types of deep learning algorithms used by your model: e.g., CNN, DNN, GAN, etc.? Does your model utilize pretrained models or transfer learning (e.g., resnet, inception, etc.)?**

The deep learning algorithm is based on a combination of inception-based convolutional neural networks and U-Net. Transfer learning techniques and regularization were also used during model training.

3. **Briefly describe the core technologies and/or framework(s) used by your model (e.g., Python, Java/JVM, TensorFlow, Pytorch, Caffe, etc.)**

Caffe and Tensorflow are used for model training and deployment.

4. **Can your model process mammography studies beyond standard 4 views (e.g., unilateral only, more than 4 views, implant)?**

The model processes 4 standard views with and without implants.

5. **How were the positive and negative labels of training images defined (e.g., specify the time interval from image to diagnosis, pathologically confirmed invasive cancer or DCIS, benign lesions included among negatives or as a third outcome, no known breast cancer diagnosis within 5 years, one or two subsequent negative screening exams)?**

The time from mammogram to diagnosis or end of follow-up was up to 4 years based on pathology confirmed invasive and in-situ cancers. Benign lesions were included for controls. The proportions of invasive, in-situ, benign lesions, normals matched a European screening population.

6. **Briefly describe your model's input requirement (e.g., DICOM, PNG, "for presentation" or "for processing" view). Is any preprocessing required?**

The model requires for presentation DICOM images and age from the image tags.

7. **Briefly describe your model's output (e.g., does it represent the probability of cancer or can it be converted to a probability?). Does the output include anything else, such as bounding boxes or lesion segmentation?**

The model outputs absolute risk of breast cancer in the population adjusted for age of the woman. The model also outputs the average risk of women at the same age.

8. How does your model generate breast-level or patient-level predictions?
Artificial intelligence analyses four images and image feature relationships between the four images. The model further uses breast cancer incidence rates and competing risks from the general population when predicting population based absolute risk of breast cancer for the woman.

9. **Does your model employ any inference-time techniques, such as model ensemble or data augmentation?**
The model uses ensemble techniques, and the model uses a time-to-event model with adjustment for competing risks.

10. Is your model able to consider any prior exams?
The model is designed for analyzing images prior to breast cancer. It analyzes images at one time point.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Vignesh Arasu
9290EED577E5445...
Author Signature

12/12/2021
Date