Long-Term Performance of an Image-Based Short-Term Risk Model for Breast Cancer

Mikael Eriksson, PhD[1]; Kamila Czene, PhD[1]; Celine Vachon, PhD[2]; Emily F. Conant, MD[3]; and Per Hall, MD, PhD[1,4]

PURPOSE Image-derived artificial intelligence–based short-term risk models for breast cancer have shown high discriminatory performance compared with traditional lifestyle/familial-based risk models. The long-term performance of image-derived risk models has not been investigated.

METHODS We performed a case-cohort study of 8,604 randomly selected women within a mammography screening cohort initiated in 2010 in Sweden for women age 40-74 years. Mammograms, age, lifestyle, and familial risk factors were collected at study entry. In all, 2,028 incident breast cancers were identified through register matching in May 2022 (206 incident breast cancers were found in the subcohort). The image-based model extracted mammographic features (density, microcalcifications, masses, and left-right breast asymmetries of these features) and age from study entry mammograms. The Tyrer-Cuzick v8 risk model incorporates self-reported lifestyle and familial risk factors and mammographic density to estimate risk. Absolute risks were estimated, and age-adjusted AUC model performances (aAUCs) were compared across the 10-year period.

RESULTS The aAUCs of the image-based risk model ranged from 0.74 (95% CI, 0.70 to 0.78) to 0.65 (95% CI, 0.63 to 0.66) for breast cancers developed 1-10 years after study entry; the corresponding Tyrer-Cuzick aAUCs were 0.62 (95% CI, 0.56 to 0.67) to 0.60 (95% CI, 0.58 to 0.61). For symptomatic cancers, the aAUCs for the image-based model were ≥0.75 during the first 3 years. Women with high and low mammographic density showed similar aAUCs. Throughout the 10-year follow-up, 20% of all women with breast cancers were deemed high-risk at study entry by the image-based risk model compared with 7.1% using the lifestyle familial-based model ($P < .01$).

CONCLUSION The image-based risk model outperformed the Tyrer-Cuzick v8 model for both short-term and long-term risk assessment and could be used to identify women who may benefit from supplemental screening and risk reduction strategies.

INTRODUCTION

Improved breast cancer risk assessment models are needed to enable personalized breast cancer screening and primary prevention.[1] Artificial intelligence (AI)–based models using data from mammograms have been reported to have higher performance than traditional lifestyle/familial-based risk models in estimating the short-term risk of breast cancer.[2-4] The long-term performance of image-derived, short-term models is yet to be determined.

Breast cancer risk assessment is currently offered in the United States and is performed using lifestyle/familial-based risk models such as Tyrer-Cuzick and Gail.[5-7] Recently, image-based risk models have emerged to refine personalized screening.[8,9] Such short-term risk tools could identify women who, after a negative or benign screening, might benefit from supplemental or more intensive screening and, may

identify women who otherwise might be diagnosed with symptomatic (breast cancer diagnosed between two screens) or later-stage breast cancer.[10]

Although mammographic screening reduces breast cancer mortality by 25%-40%,[11,12] the incidence of breast cancer has steadily increased in the Western world over the past 50 years.[13] Currently, approximately 13% of women are diagnosed with breast cancer in their lifetime. Prophylactic interventions, including lifestyle changes and risk-reducing medications, have shown promising results in reducing breast cancer incidence in women with an increased risk of breast cancer.[14] For women determined to have high risk on the basis of the 10-year or lifetime risk of breast cancer, the National Institute for Health and Care Excellence (NICE) and US Preventive Services Task Force (USPSTF) clinical guidelines recommend risk-reducing interventions or more intense screening.[15,16]

CONTEXT

Key Objective

We compared an image-based risk model with a clinical lifestyle/familial-based risk model for predicting breast cancer risk in mammography screening throughout the 10-year follow-up.

Knowledge Generated

In this case-cohort study, 8,604 women age 40-74 years were randomly selected from a Swedish screening cohort which began in 2010 together with 2,028 incident breast cancers diagnosed before May 2022. The image-based model showed higher discriminatory performance compared with the clinical model for women overall throughout the 10-year follow-up and for symptomatic breast cancers in the short-term and estrogen-positive, screen detected, and invasive breast cancers in the long-term.

Relevance *(G. Fleming)*

We should work toward incorporation of image-based models to identify those patients at higher risk for development of breast cancer who may benefit from enhanced screening or prevention strategies.*

*Relevance section written by *JCO* Associate Editor Gini Fleming, MD.

Using the unique KARolinska MAmmography Project for Risk Prediction of Breast Cancer (KARMA) screening cohort,[17] we investigated the long-term and short-term performances of the image-based risk model and compared the results with those of the Tyrer-Cuzick v8 risk model. In our study, we used an independent set of baseline mammograms from the same underlying screening cohort that was used to develop the model.[3] The image-based risk model is a 2-year risk model designed for women attending biennial screening and use image-derived data to predict the risk of being diagnosed with a symptomatic cancer or a cancer at the next routine screening.[10] The model is available for clinical use in the United States and Europe. The Tyrer-Cuzick model is an established risk tool on the basis of lifestyle, familial risk factors, and mammographic density to identify women who may benefit from intervention or follow-up.[16,18,19] We estimated the discriminatory performances of the two models for predicting breast cancer diagnoses for up to 10 years after study entry for potential use in primary prevention. In addition, we investigated both models for assessing short-term risk to identify women who could benefit from supplemental screening or shorter screening intervals. We performed analyses for subtype-specific breast cancers and estimated the potential impact of how masking of a tumor by mammographic dense tissue may affect model performance. Finally, we compared the ability of the two models to correctly stratify women according to their risk of breast cancer.

METHODS

Study Population

Sweden has a national mammography screening program in which women age 40-74 years are invited every 18 or 24 months for screening, depending on age and region.[20] Women who underwent mammographic screening between October 2010 and March 2013 were invited to enter the cohort in the prospective KARMA study.[17] The approximately 70,000 women who chose to participate consented in writing to participate in studies on the risk and prognosis of breast cancer. Participants responded to a web survey on lifestyle and familial-related breast cancer risk factors. In addition, women donated blood, accepted linkage to national breast cancer registers, and allowed long-term storage of their mammograms for image analysis. Women who were diagnosed with breast cancer after 3 months of a negative screen were eligible for our study. We performed a case-cohort study that included a random sample of 8,604 women (12.9%) from the 66,814 eligible women in the KARMA cohort. The case-cohort also included all incident breast cancers that were diagnosed in the underlying cohort at the time of register linkage.

The study was approved by the Swedish Ethical Review Authority (2010/958-31/1).

Risk Factors at Study Entry and Follow-Up of Breast Cancers

Full-field digital mammograms at study entry from left and right breasts in mediolateral oblique and craniocaudal views were used to derive AI-assessed mammographic features using the ProFound AI Risk (Nashua, NH) tool and Stratus density tool as previously described.[3,21] The details on risk factor definitions are presented in the Data Supplement (online only). The absolute risk of breast cancer was calculated on the basis of age, mammographic features (density, microcalcifications, and masses), left-right breast asymmetries of these features, risk factor distributions, and national statistics on breast cancer incidence rate and competing mortality risk.

Self-reported lifestyle and familial risk factors were used to calculate the absolute risk of breast cancer using the

Tyrer-Cuzick model v8.[18] We added the Tyrer-Cuzick tool to our study to allow comparison with a risk assessment tool often used in other countries for risk assessment.[16,19] The tool calculates breast cancer risk on the basis of lifestyle and familial risk factors, that is, age, height, weight, age at menarche, age at first childbirth, menopausal status, use of hormone replacement therapy, benign breast disorders, first- and second-degree family history of breast cancer, first-degree family history of ovarian cancer, and mammographic density.[21]

Breast cancer status and mode of detection were retrieved for breast cancers diagnosed up to May 2022 from the national quality breast cancer registry through linkage using Swedish personal identification numbers.[22] Interval cancer was defined as a cancer diagnosed between two screens. Breast cancer status was also cross-checked through linkage to the national cause-of-death register at the end of follow-up for women who died during the study period.[23]

Statistical Analysis

Descriptive statistics reported study participant characteristics and distribution of risk factors at study entry.[24] The time that passed from the date of mammogram at study entry to the date of breast cancer diagnosis was reported as the frequency distribution. Absolute risks were estimated at study entry using the image-based risk model and Tyrer-Cuzick model after 1-10 years of follow-up. Model performances for the two models were estimated using odds ratio per adjusted standard deviation (OPERA) and using AUC after adjustment for age (aAUC)[25-27] (Data Supplement). Inverse probability weights were used to account for the case-cohort sampling.[24,28] The 10-year risk assessment included all women in the subcohort and the incident breast cancers that were diagnosed after 3 months to 10 years after study entry. When estimating the risk performance for a defined year of follow-up, we included breast cancers that were diagnosed after 3 months of study entry up to that specific year. The risk model performances were estimated at the end of each year of follow-up. The 95% CIs of the aAUC point estimates were estimated using 1,000 bootstraps.[27] We estimated the significance of the differences between aAUCs by comparing the two risk models using bootstrapping.[29] The discriminatory performance was reported for the full study population and additionally for subgroups of women by family history of breast cancer, postmenopausal status, estrogen receptor status, mode of detection, and women with above/below median mammographic density to assess the likelihood of cancer masking by dense breast tissue. Tyrer-Cuzick risk scores were missing because of incomplete survey responses in a subset of 233 women (124 cases and 109 noncases). For this reason, we performed a sensitivity analysis of the aAUCs by comparing the image-based risk and Tyrer-Cuzick risk models in women with complete data. The proportions of women at low, general, moderate, and high risk on the basis of 2-year

and 10-year risk scores were reported for the two risk models using the NICE and USPSTF guidelines. Positive predictive values (PPVs) were estimated with 95% CIs.[30]

Statistical analyses were performed using R 4.1.[31] All tests were two-sided at a significance level of 0.05.

RESULTS

Study Population

The case-cohort consisted of 8,604 women in the subcohort and 2,028 incident breast cancers identified at the 5/2022 registry linkage (including 206 cases in the subcohort) that were diagnosed after 3 months to 10 years after study entry, Table 1. At baseline, the mean age was 56.62 (±9.50) years for cases and 53.90 (±9.87) years for the subcohort. BMI was 24.88 and 25.25 for cases and for the subcohort, respectively. Sixty-seven percent of cases and 55% of women in the subcohort were postmenopausal. Family history of breast cancer was reported in 21% of cases and 12% of the subcohort. Benign breast disease was more common in the cases (31%) than in the subcohort (22%). The mean absolute 2-year risk score at study entry was 0.63% and 0.60% by the image-based model and the Tyrer-Cuzick model, respectively, in the subcohort. In cases, the corresponding 2-year risk scores were 1.21% and 0.79%. The absolute 10-year Tyrer-Cuzick risk score was approximately five-fold higher than the corresponding Tyrer-Cuzick 2-year risks.

Breast Cancer and Follow-Up Time from Study Entry

The follow-up period from initial mammogram to breast cancer diagnosis ranged from >3 months to 10 years and, half of the events occurred within the first 5 years (Data Supplement). After the 10-years follow-up, 62% of the breast cancers were screen detected and 38% were nonscreen detected; 86% of the cancers were estrogen receptor (ER)–positive and 14% ER-negative; 86% of the cancers were invasive and 14% in situ (Data Supplement).

Predictive Performance

The discriminatory performance of the image-based risk model ranged from an aAUC of 0.74 (95% CI, 0.70 to 0.78) after 1-year of follow-up to an aAUC of 0.65 (95% CI, 0.63 to 0.66) after 10 years of follow-up, Table 2, Figure 1. The corresponding 1- and 10-year aAUCs for premenopausal women were 0.69 (95% CI, 0.62 to 0.75) and 0.62 (95% CI, 0.60 to 0.65) and for postmenopausal women 0.76 (95% CI, 0.69 to 0.82) and 0.66 (95% CI, 0.64 to 0.67), respectively. The 1- and 10-year aAUCs in women stratified by breast density above/below median were 0.69 (95% CI, 0.63 to 0.75) and 0.75 (95% CI, 0.65 to 0.83) after 1 year and 0.64 (95% CI, 0.63 to 0.66) and 0.64 (95% CI, 0.62 to 0.66) after 10 years. The aAUCs for women without FH ranged from 0.75 (95% CI, 0.70 to 0.80) after 1 year to 0.66 (95% CI, 0.64 to 0.67) after 10 years,

TABLE 1. Study Participant Characteristics at Study Entry Including (N = 10,426) Study Participants With up to 10 Years of Follow-Up

Characteristic	Subcohort[a] (N = 8,604)	Cases[a] (N = 2,028)	P[b]
Age, years	53.90 (9.87)	56.62 (9.50)	<.01
BMI, mean (SD)	25.25 (4.33)	24.88 (5.62)	.30
Postmenopausal, n/N (%)	4,706/8,604 (55)	1,366/2,028 (67)	<.01
Family history of breast cancer, n/N (%)			<.01
No family history	7,229/8,237 (88)	1,507/1,896 (79)	
First-degree relative age <50 years	261/8,237 (3.2)	96/1,896 (5.1)	
First-degree relative age ≥50 years	747/8,237 (9.1)	293/1,896 (15)	
Breast cancer in 2nd/3rd-degree relative, n/N (%)	762/5,723 (13)	189/1,216 (16)	.04
Age at menarche, years, mean (SD)	13.06 (1.48)	13.11 (1.45)	.17
Age at first childbirth, years, mean (SD)	27.17 (5.29)	27.12 (5.20)	.76
Current use of HRT,[c] n/N (%)	319/8,103 (3.9)	124/1,849 (6.7)	<.01
Benign breast disease, n/N (%)	1,865/8,378 (22)	591/1,878 (31)	<.01
Percent mammographic density, mean (SD)	25.41 (19.43)	28.10 (19.45)	<.01
Mammographic density above median, n/N (%)	4,213/8,604 (49)	1,108/2,028 (55)	<.01
Image risk model, 2-year risk score, mean (SD)	0.63 (0.96)	1.21 (1.86)	<.01
Tyrer-Cuzick, 2-year risk score, mean (SD)	0.60 (0.40)	0.79 (0.53)	<.01
Tyrer-Cuzick 10-year risk score, mean (SD)	3.08 (1.86)	3.94 (2.50)	<.01

Abbreviations: HRT, hormone replacement therapy; SD, standard deviation.

[a]Subcohort of noncases and the 206 incident cancers diagnosed in the subcohort. For cases, all incident cases outside and within the subcohort are included.

[b]Wilcoxon rank sum test; Pearson's χ^2 test.

[c]Hormone replacement therapy.

whereas the corresponding aAUCs for women with FH were 0.65 (95% CI, 0.53 to 0.75) and 0.62 (95% CI, 0.59 to 0.65).

The Tyrer-Cuzick risk performance aAUCs were 0.62 (95% CI, 0.56 to 0.67) and 0.60 (95% CI, 0.58 to 0.61) after 1 and 10 years of follow-up, respectively (Table 2, Fig 1). The aAUCs for women without family history ranged from 0.61 (95% CI, 0.53 to 0.67) after 1 year to 0.58 (95% CI, 0.56 to 0.59) after 10 years of follow-up. The corresponding aAUCs for women with family history were significantly higher of 0.52 (95% CI, 0.39 to 0.64) and 0.56 (95% CI, 0.53 to 0.60). Similar performances were observed in a subgroup of 1,904 breast cancers and 8,483 women in the subcohort where the risk scores were available for both risk models (Data Supplement).

When comparing aAUCs of the image-based risk model to corresponding Tyrer-Cuzick aAUCs at 1, 2, 5, and 10 years of follow-up, the image-based risk model estimates showed significantly higher aAUCs across the follow-up periods in range 0.12 ($P < .01$) to 0.05 ($P < .01$; Fig 1). The image-based risk model also showed significantly higher aAUC in women with high mammographic density and in women with no family history in the range of 1-10 years of follow-up (Table 2, Data Supplement). A detailed description of OPERA performances is provided in the Data Supplement.

Mode of Detection, Estrogen Receptor Status, and Tumor Invasiveness

The discriminatory performances of the image-based risk model in the short term showed aAUCs of ≥0.70 for all breast cancer subtypes within 1-3 years of follow-up (Fig 2). Symptomatic and ER-negative breast cancers showed the highest point estimates in the short-term, aAUCs ≥0.75. In the long-term, all subtypes showed aAUCs of ≥0.62 after 10 years of follow-up. Compared with the image-based model, the Tyrer-Cuzick model showed significantly lower aAUCs of ≤0.65 for symptomatic breast cancers in the short-term and, lower aAUCs of ≤0.62 for ER-positive, screen detected, and invasive breast cancers in the long-term ($P < .01$) after adjustment for multiple comparison (Data Supplement).

Risk Stratification and PPV

The proportion of breast cancers identified as high-risk using the image-based risk model (20%) was significantly larger than the corresponding proportion identified using Tyrer-Cuzick (7.1%, $P < .01$; Fig 3). However, the image-based risk model categorized 6.3% of the subcohort as high-risk; the corresponding number for the Tyrer-Cuzick model was 2.4%. Using the NICE guidelines, image-based risk showed approximately 20-fold risk stratification comparing the women at high risk with the

TABLE 2. Discriminatory Performance After Adjustment for Age of the Image-Based Risk Model and the Tyrer-Cuzick v8 Risk Model in Subgroups of Women by Menopausal Status, Mammographic Density, and Family History of Breast Cancer at Study Entry

Group of Women	1 Year, AUC (95% CI)	2 Years, AUC (95% CI)	3 Years, AUC (95% CI)	4 Years, AUC (95% CI)	5 Years, AUC (95% CI)	6 Years, AUC (95% CI)	8 Years, AUC (95% CI)	10 Years, AUC (95% CI)
Image-based model								
All women	0.74 (0.70 to 0.78)	0.73 (0.70 to 0.75)	0.72 (0.70 to 0.74)	0.70 (0.68 to 0.72)	0.69 (0.67 to 0.71)	0.68 (0.66 to 0.69)	0.65 (0.64 to 0.67)	0.65 (0.63 to 0.66)
Postmenopausal	0.76 (0.69 to 0.82)	0.72 (0.69 to 0.75)	0.72 (0.69 to 0.75)	0.70 (0.68 to 0.73)	0.69 (0.67 to 0.71)	0.68 (0.66 to 0.70)	0.66 (0.64 to 0.68)	0.66 (0.64 to 0.67)
Premenopausal	0.69 (0.62 to 0.75)	0.73 (0.69 to 0.78)	0.72 (0.67 to 0.76)	0.68 (0.65 to 0.72)	0.68 (0.65 to 0.71)	0.66 (0.63 to 0.69)	0.63 (0.60 to 0.65)	0.62 (0.60 to 0.65)
Low breast density[a]	0.75 (0.65 to 0.83)	0.72 (0.68 to 0.75)	0.71 (0.67 to 0.74)	0.68 (0.65 to 0.71)	0.67 (0.64 to 0.70)	0.66 (0.64 to 0.69)	0.64 (0.62 to 0.67)	0.64 (0.62 to 0.66)
High breast density[a]	0.69 (0.63 to 0.75)	0.73 (0.70 to 0.76)	0.73 (0.69 to 0.76)	0.70 (0.67 to 0.72)	0.69 (0.67 to 0.72)	0.68 (0.66 to 0.70)	0.65 (0.63 to 0.67)	0.64 (0.63 to 0.66)
No family history[b]	0.75 (0.70 to 0.80)	0.74 (0.71 to 0.77)	0.74 (0.71 to 0.77)	0.72 (0.69 to 0.74)	0.70 (0.68 to 0.72)	0.69 (0.67 to 0.71)	0.66 (0.65 to 0.68)	0.66 (0.64 to 0.67)
Family history[b]	0.65 (0.53 to 0.75)	0.67 (0.62 to 0.73)	0.67 (0.62 to 0.72)	0.65 (0.60 to 0.69)	0.65 (0.61 to 0.68)	0.63 (0.59 to 0.67)	0.62 (0.58 to 0.65)	0.62 (0.59 to 0.65)
Tyrer-Cuzick[c]								
All women	0.62 (0.56 to 0.67)	0.62 (0.59 to 0.65)	0.62 (0.59 to 0.64)	0.62 (0.59 to 0.64)	0.60 (0.58 to 0.62)	0.60 (0.58 to 0.62)	0.59 (0.58 to 0.61)	0.60 (0.58 to 0.61)
Postmenopausal	0.66 (0.58 to 0.75)	0.64 (0.61 to 0.68)	0.63 (0.6 to 0.66)	0.64 (0.61 to 0.66)	0.62 (0.60 to 0.64)	0.63 (0.60 to 0.65)	0.61 (0.59 to 0.63)	0.62 (0.60 to 0.63)
Premenopausal	0.61 (0.52 to 0.68)	0.58 (0.52 to 0.65)	0.60 (0.54 to 0.65)	0.58 (0.54 to 0.62)	0.57 (0.53 to 0.61)	0.56 (0.53 to 0.60)	0.55 (0.53 to 0.58)	0.56 (0.54 to 0.59)
Low breast density[a]	0.64 (0.52 to 0.74)	0.64 (0.59 to 0.68)	0.63 (0.59 to 0.66)	0.62 (0.59 to 0.65)	0.61 (0.58 to 0.64)	0.61 (0.58 to 0.63)	0.59 (0.57 to 0.62)	0.60 (0.58 to 0.62)
High breast density[a]	0.58 (0.50 to 0.65)	0.58 (0.53 to 0.62)	0.57 (0.53 to 0.61)	0.57 (0.54 to 0.60)	0.56 (0.54 to 0.59)	0.56 (0.54 to 0.59)	0.56 (0.53 to 0.58)	0.56 (0.54 to 0.58)
No family history[b]	0.61 (0.53 to 0.67)	0.60 (0.57 to 0.64)	0.59 (0.56 to 0.62)	0.60 (0.57 to 0.62)	0.58 (0.56 to 0.60)	0.58 (0.56 to 0.60)	0.57 (0.55 to 0.59)	0.58 (0.56 to 0.59)
Family history[b]	0.52 (0.39 to 0.64)	0.58 (0.52 to 0.64)	0.59 (0.54 to 0.64)	0.59 (0.54 to 0.63)	0.56 (0.52 to 0.61)	0.56 (0.52 to 0.60)	0.55 (0.52 to 0.59)	0.56 (0.53 to 0.60)

NOTE. Multiple estimations were performed for breast cancers diagnosed within 10 years after study entry.

[a]Low/high mammographic density was defined as below/above median mammographic density in the study sample.

[b]First-degree relative with breast cancer (mother, sister, and child).
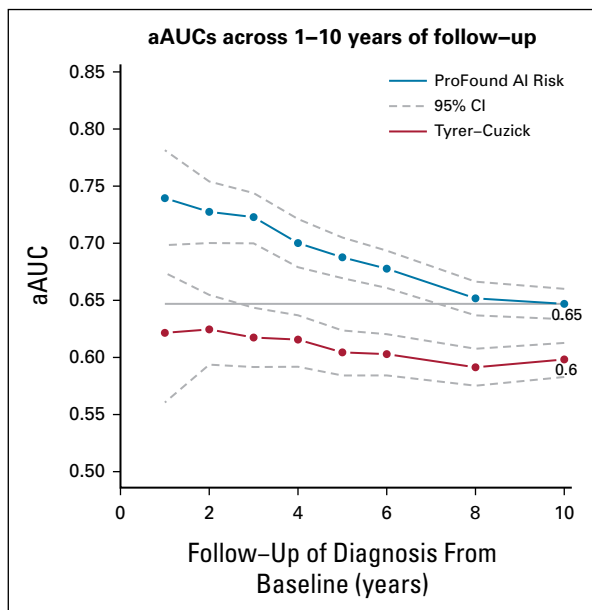
[c]Tyrer-Cuzick v8.

FIG 1. Discriminatory performance after adjustment for age of the image-based risk model and the Tyrer-Cuzick v8 risk model throughout the first years after study entry. The image-based risk model included 8,604 women in the subcohort and 2,208 breast cancers. Tyrer-Cuzick included 8,483 women in the subcohort and 1,904 breast cancers. The horizontal line denotes the aAUC point estimate (0.65) of the image-based risk model for 10-year of follow-up in the study sample. The corresponding aAUC for Tyrer-Cuzick was 0.60. The aAUC point estimate differences between the image-based risk model and Tyrer-Cuzick ranged from 0.12 (1-year follow-up, $P < .01$) to 0.05 (10-year follow-up, $P < .01$). aAUC, AUC after adjustment for age; AI, artificial intelligence.

women at low risk. In comparison, the corresponding risk stratification for Tyrer-Cuzick was approximately 14-fold. A similar risk stratification was observed when 2-year risk was predicted using Tyrer-Cuzick (Data Supplement). When comparing high-risk women with women at general risk using the USPSTF risk categorization, the image-based risk stratification was approximately 11-fold compared with approximately 7-fold for Tyrer-Cuzick (Data Supplement).

The PPVs ranged from 5.4% (95% CI, 5.2 to 5.6) to 5.5% (95% CI, 5.3 to 5.7) across 1%-20% of women at the highest risk using the image-based risk model (Data Supplement). The corresponding PPVs for Tyrer-Cuzick were significantly lower ranging from 4.4 (95% CI, 4.2 to 4.5) to 5.0 (95% CI, 4.8 to 5.1), $P < .01$.

DISCUSSION

We investigated the performance of an image-based AI short-term risk model in a 10-year follow-up study and compared it with a traditional, lifestyle/familial-based risk model. The image-based risk model showed significantly better discriminatory performance in the 10-year follow-up period with a mean follow-up time of 5 years. In addition,

the image-based model showed a significant advantage over the lifestyle/familial-based model in the short-term for women with symptomatic breast cancers and in the long-term for women with ER-positive cancer, screen detected, and invasive breast cancers.

Image-based models have shown to perform well in the short term and are designed to improve early detection compared with standard-of-care screening practices.[2,3,9] Short-term risk models allow for identifying women who may benefit from supplemental screening or shorter screening intervals to identify breast cancers earlier. If the time to diagnosis is shortened, outcomes could improve significantly affecting health economy because of decreases in treatment costs.[32,33]

At the same time, long-term risk assessment is needed in primary prevention of breast cancer.[34,35] Considering that the tumor development time for screen or clinically detected cancer is estimated to be 5-20 years,[36] a 10-year or lifetime risk projection is reasonable. Guidelines for primary prevention are also available to recommend lifestyle changes and medical intervention.[14,37] Novel approaches have also been suggested where endocrine medical interventions could reduce mammographic dense tissue and improve screening sensitivity and, therefore, outcomes.[38]

Clinical guidelines such as the NICE and USPSTF support the use of lifestyle/familial-based and genetic risk models for supplemental screening and risk-reducing intervention.[16,19] There may be a significant performance gain with the incorporation of image-based data beyond breast density to refine risk assessment, and investigations should, therefore, be conducted to expand guidelines to incorporate such newer, image-based risk tools.

In the United States, approximately 12% of women screened have a lifetime risk of ≥20% and are eligible for supplemental screening with breast MRI.[39] Approximately 39% of cancers could be captured by identifying the high-risk group using the Tyrer-Cuzick risk model.[40] On the basis of the 10-year follow-up in our Swedish study, 2.4% of the women were at high short-term risk using Tyrer-Cuzick capturing 7.1% of the breast cancers. In comparison, the image-based risk model using data beyond just breast density identified 6.3% of women at high risk, capturing 20% of breast cancers. In Sweden, risk assessment is not performed for the screening population. When considering a risk-stratified screening approach, potential risks for adding clinical follow-up and potentially, additional interventions at a large scale require further investigation.

Although the image-based risk model predicts a higher proportion of breast cancers compared with the lifestyle/familial-based risk model, it could be that subgroups of women are captured by the lifestyle/familial-based model only. We showed that when stratifying women on family history of breast cancer, menopausal status, women with high versus low density, and breast cancer ER
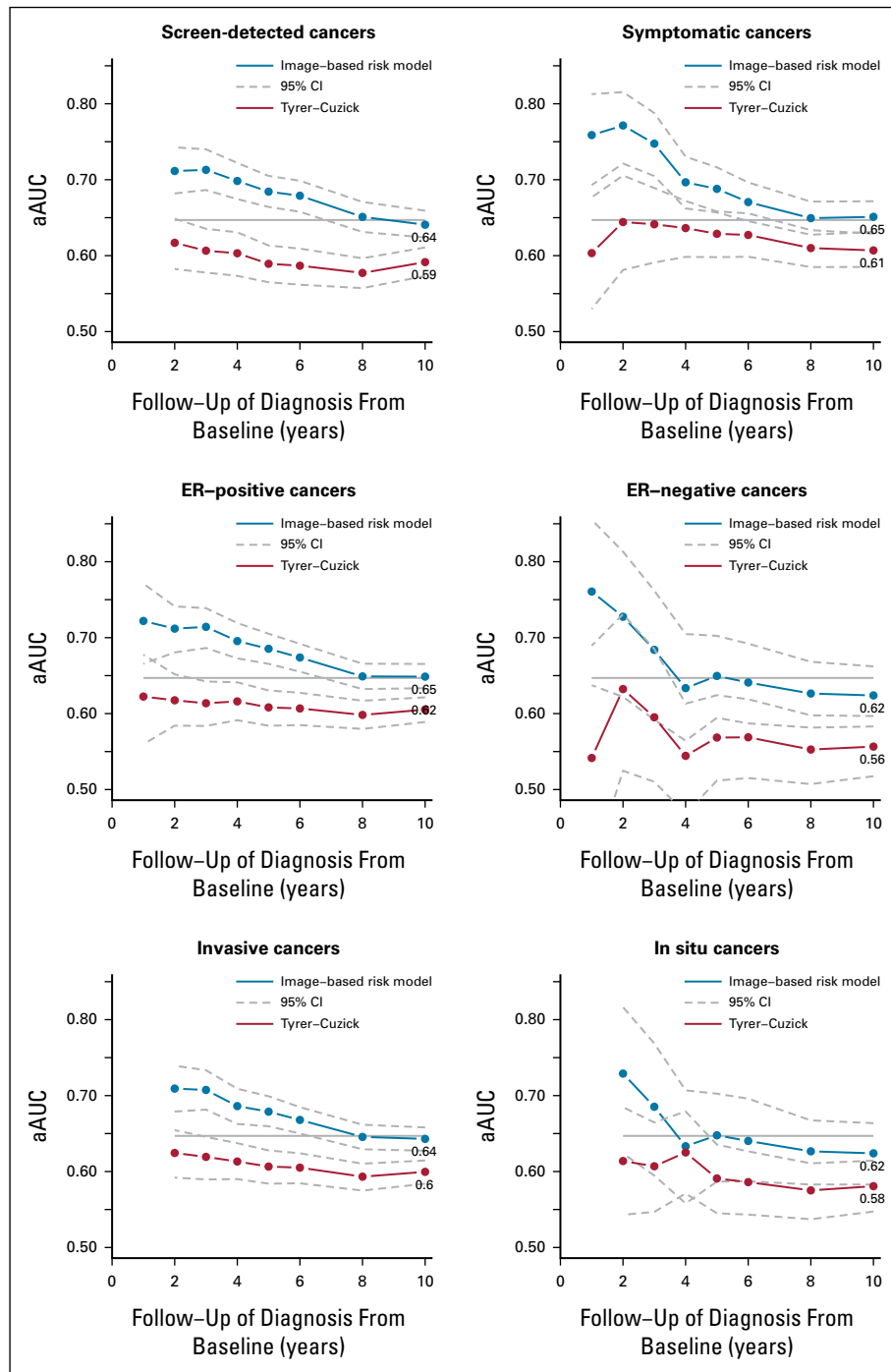
FIG 2. Discriminatory performance after adjustment for age of the image-based risk model and the Tyrer-Cuzick v8 risk model at study entry stratified by mode of detection, estrogen receptor status, and tumor invasiveness. Multiple estimations were performed for breast cancers diagnosed within 10 years after study entry. The horizontal line denotes the aAUC point estimate (0.65) of the image-based risk model for 10-year follow-up in the full study sample. The aAUCs for the image-based model were significantly higher after 10 years of follow-up for screen detected, symptomatic, ER-positive, and invasive breast cancers, $P < .01$. aAUC, AUC after adjustment for age; ER, estrogen receptor.

status, mode of detection, and invasiveness, the image-based approach showed significantly higher or similar discriminatory performance compared with the lifestyle/ familial-based risk assessment approach throughout the 10-year follow-up. We observed a reduced discriminatory performance in women with a family history of breast cancer

Risk[a]	Cases, No. (%)	Subcohort, No. (%)	Absolute risk, %[b]	RR[c]
Image-based risk, 2-year				
Low	84 (4.1)	782 (9.1)	0.12	0.36
General	875 (43)	5,301 (62)	0.33	1.0 (ref.)
Moderate	670 (33)	1,980 (23)	0.83	2.5
High	399 (20)	541 (6.3)	2.33	7.1
Tyrer-Cuzick, 10-year				
Low	1 (< 0.1)	21 (0.2)	0.69	0.34
General	870 (46)	5,254 (62)	2.05	1 (ref.)
Moderate	898 (47)	3,004 (35)	4.02	2.0
High	135 (7.1)	204 (2.4)	9.74	4.7
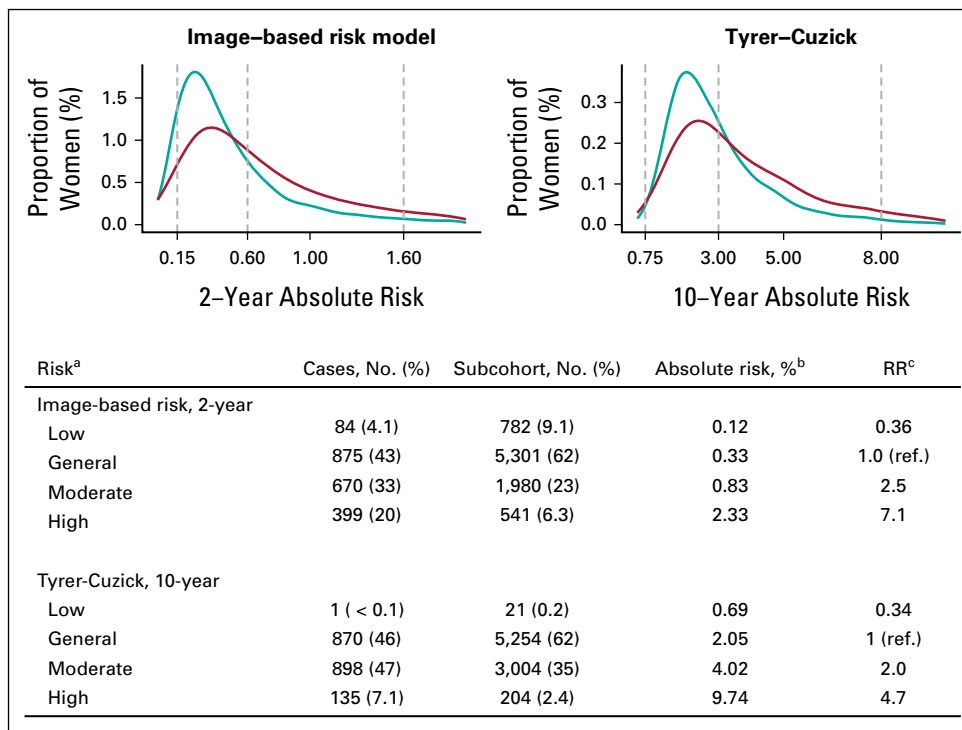
FIG 3. Frequency distribution of absolute risks at study entry for 2,028 breast cancers (red) and 8,604 women in the subcohort (teal) using the image-based risk model and Tyrer-Cuzick v8 models and categorization of women into high, moderate, general, and low risk following NICE guidelines in relation to breast cancers diagnosed in the 10-year follow-up period. [a]The NICE guidelines 10-year absolute risk categories were general, moderate, and high using absolute risk cutoffs 3% and 8%. A low-risk category was added using cutoff 0.75%. For the imaging-only AI risk model and the expanded AI risk model, the risk cutoffs were adapted to 2-year risks by dividing the 10-year risk cutoffs by 5. This resulted in the cutoffs 0.15%, 0.6%, and 1.6%. [b]Median absolute risks of the subcohort are presented using the 2-year risk scale for the two AI-based models and the 10-year risk scale for Tyrer-Cuzick. [c]The risk ratio was calculated as ratios of the absolute risks in each absolute risk category. Test was performed for the difference between the binomial proportions of breast cancers captured at high risk by the image-based risk model and the Tyrer-Cuzick risk model. The proportion of cancer captured by the image-based risk model (20%) was significantly higher compared with Tyrer-Cuzick (7.1%), $P < .01$. The risk ratio between high-risk and low-risk was 19.9 for the image-based risk model and 14.0 for Tyrer-Cuzick. Tyrer-Cuzick risk scores were missing for n = 124 cases and n = 109 noncases. AI, artificial intelligence; RR, risk ratio; NICE, National Institute for Health and Care Excellence.

for the image-based and Tyrer-Cuzick risk models, which warrants further investigation.

We observed that the image-based risk tool performed with similar risk stratification performance in women with low and high mammographic densities. In standard of care, screening sensitivity for women with extremely dense breasts is approximately 50% compared with approximately 90% in women with almost entirely fatty breasts.[41,42] Therefore, the challenging radiological task of detecting a tumor masked by dense tissue could be reduced by using an image-based risk tool for identifying women who benefit from supplemental screening following a negative or benign screen.[43]

The overall discriminatory performance decreased across the 1-10-year follow-up by 9% and 2% using image-based

and Tyrer-Cuzick risk models, respectively, resulted in the image-based model demonstrating a higher performance in the range of 12% to 5%.

Our study has several limitations. We investigated the model performance in a Swedish large-scale prospective screening population where the vast majority of women were White, attending biennial screening using 2D mammography (GE, Philips, Sectra, Siemens, Fuji machines) with no additional supplemental screening, a recall rate of approximately 3%, and a cancer detection rate of approximately 5/1,000 examinations.[44] Further studies are needed to investigate the external validity of our findings in other screening settings. Our effect estimates using the Tyrer-Cuzick model were affected by missing risk exposure data, including information on

2nd and 3rd relatives with breast cancer. This was likely nondifferential between cases and noncases in this prospective data collection. In our study, we examined the image-based only model of our risk model.[3] Further studies are needed to study the addition of lifestyle and familial risk factors, as well as genetic determinants to the image-based risk model.

In conclusion, an image-based risk model developed for short-term risk to identify women who may benefit from supplemental screening can also be used to assess long-term risk for identifying women who may benefit from primary prevention. The image-based model showed better discriminatory accuracy for long-term risk assessment than the clinical Tyrer-Cuzick v8 model.

AFFILIATIONS

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[2]Mayo Clinic College of Medicine, Rochester, MN
[3]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
[4]Department of Oncology, Södersjukhuset University Hospital, Stockholm, Sweden

CORRESPONDING AUTHOR

Mikael Eriksson, PhD, Department of Medical Epidemiology and Biostatistics, Karolinska institutet, Nobels vag 12A, 171 77 Stockholm, Sweden; e-mail: mikael.eriksson@ki.se.

DISCLAIMER

The funders had no role in the study design, management, analyses, interpretation of data, writing the manuscript, approval, or decision to publish the results.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at DOI https://doi.org/10.1200/JCO.22.01564.

DATA SHARING STATEMENT

The data sets for this study fall under the European Union General Data Protection Regulation (GDPR) legislation and are available on reasonable request.

AUTHOR CONTRIBUTIONS

Conception and design: Mikael Eriksson, Kamila Czene, Per Hall
Financial support: Mikael Eriksson, Per Hall
Administrative support: Per Hall
Provision of study materials or patients: Emily F. Conant, Per Hall
Collection and assembly of data: Mikael Eriksson, Emily F. Conant, Per Hall
Data analysis and interpretation: Mikael Eriksson, Kamila Czene, Celine Vachon, Emily F. Conant, Per Hall
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

REFERENCES

1. Gail MH, Pfeiffer RM: Breast cancer risk model requirements for counseling, prevention, and screening. J Natl Cancer Inst 110:994-1002, 2018

2. Yala A, Mikhael PG, Strand F, et al: Toward robust mammography-based models for breast cancer risk. Sci Transl Med 13:eaba4373, 2021

3. Eriksson M, Czene K, Strand F, et al: Identification of women at high risk of breast cancer who need supplemental screening. Radiology 297:327-333, 2020

4. Louro J, Posso M, Hilton Boon M, et al: A systematic review and quality assessment of individualised breast cancer risk prediction models. Br J Cancer 121:76-85, 2019

5. Smith RA, Andrews KS, Brooks D, et al: Cancer screening in the United States, 2019: A review of current American cancer society guidelines and current issues in cancer screening. CA: A Cancer J Clin 69:184-210, 2019

6. Bevers TB, Ward JH, Arun BK, et al: Breast cancer risk reduction, version 2.2015. J Natl Compr Canc Netw 13:880-915, 2015

7. Monticciolo DL, Newell MS, Hendrick RE, et al: Breast cancer screening for average-risk women: Recommendations from the ACR commission on breast imaging. J Am Coll Radiol 14:1137-1143, 2017

8. Gastounioti A, Desai S, Ahluwalia VS, et al: Artificial intelligence in mammographic phenotyping of breast cancer risk: A narrative review. Breast Cancer Res 24:14, 2022

9. McKinney SM, Sieniek M, Godbole V, et al: International evaluation of an AI system for breast cancer screening. Nature 577:89-94, 2020

10. Eriksson M, Czene K, Pawitan Y, et al: A clinical model for identifying the short-term risk of breast cancer. Breast Cancer Res 19:29, 2017

11. Lauby-Secretan B, Scoccianti C, Loomis D, et al: Breast-cancer screening—Viewpoint of the IARC working group. N Engl J Med 372:2353-2358, 2015

12. Tabar L, Chen THH, Yen AMF, et al: Early detection of breast cancer rectifies inequality of breast cancer outcomes. J Med Screen 28:34-38, 2021

13. Chen Z, Xu L, Shi W, et al: Trends of female and male breast cancer incidence at the global, regional, and national levels, 1990–2017. Breast Cancer Res Treat 180:481-490, 2020

14. Mocellin S, Goodwin A, Pasquali S: Risk-reducing medications for primary breast cancer: A network meta-analysis. Cochrane Database Syst Rev 4:CD012191, 2019

15. National Collaborating Centre for Cancer: Familial Breast Cancer: Classification and Care of People at Risk of Familial Breast Cancer and Management of Breast Cancer and Related Risks in People with a Family History of Breast Cancer. Cardiff, UK, NCCC, 2013

16. Siu AL, UPST Force: Screening for breast cancer: US Preventive Services Task Force Recommendation Statement. Ann Intern Med 164:279-296, 2016

17. Gabrielson M, Eriksson M, Hammarstrom M, et al: Cohort profile: The Karolinska mammography Project for risk prediction of breast cancer (KARMA). Int J Epidemiol 46:1740-1741g, 2017

18. Tyrer J, Duffy SW, Cuzick J: A breast cancer prediction model incorporating familial and personal risk factors. Stat Med 23:1111-1130, 2004

19. National Institute for Health and Care Excellence: Familial Breast Cancer: Classification, Care and Managing Breast Cancer and Related Risks in People with a Family History of Breast Cancer, London, UK, NICE, 2013

20. Zidar MN, Larm P, Tillgren P, et al: Non-attendance of mammographic screening: The roles of age and municipality in a population-based Swedish sample. Int J Equity Health 14:157, 2015

21. Eriksson M, Li J, Leifland K, et al: A comprehensive tool for measuring mammographic density changes over time. Breast Cancer Res Treat 169:371-379, 2018

22. Ehinger A, Lindman H, Löfgren L, et al: Årsrapport 2015 från Nationella Bröstcancerregistret, 2016

23. Brooke HL, Talbäck M, Hörnblad J, et al: The Swedish cause of death register. Eur J Epidemiol 32:765-773, 2017

24. Sharp SJ, Poulaliou M, Thompson SG, et al: A review of published analyses of case-cohort studies and recommendations for future reporting. PLoS One 9: e101176, 2014

25. Hopper JL: Odds per adjusted standard deviation: Comparing strengths of associations for risk factors measured on different scales and across diseases and populations: Table 1. Am J Epidemiol 182:863-867, 2015

26. Janes H, Pepe MS: Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. Biometrika 96: 371-382, 2009

27. de Carvalho VI, Rodriguez-Alvarez MX: Bayesian Nonparametric Inference for the Covariate-Adjusted ROC Curve. arXiv preprint arXiv:1806.00473, 2018

28. Barlow WE, Ichikawa L, Rosner D, et al: Analysis of case-cohort designs. J Clin Epidemiol 52:1165-1172, 1999

29. Carpenter J, Bithell J: Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. Stat Med 19:1141-1164, 2000

30. Altman DG, Bland JM: Statistics notes: Diagnostic tests 2: Predictive values. BMJ 309:102, 1994

31. RC Team: A Language and Environment for Statistical Computing. Vienna, Austria, RC Team, 2013

32. Consedine NS, Magai C, Krivoshekova YS, et al: Fear, anxiety, worry, and breast cancer screening behavior: A critical review. Cancer Epidemiol Biomarkers Prev 13:501-510, 2004

33. Schousboe JT, Kerlikowske K, Loh A, et al: Personalizing mammography by breast density and other risk factors for breast cancer: Analysis of health benefits and cost-effectiveness. Ann Intern Med 155:10-20, 2011

34. Smith SG, Sestak I, Forster A, et al: Factors affecting uptake and adherence to breast cancer chemoprevention: A systematic review and meta-analysis. Ann Oncol 27:575-590, 2016

35. Chlebowski RT: Reducing the risk of breast cancer. N Engl J Med 343:191-198, 2000

36. Ryu EB, Chang JM, Seo M, et al: Tumour volume doubling time of molecular breast cancer subtypes assessed by serial breast ultrasound. Eur Radiol 24: 2227-2235, 2014

37. Owens DK, Davidson KW, Krist AH, et al: Medication use to reduce risk of breast cancer: US Preventive Services Task Force recommendation statement. Jama 322:857-867, 2019

38. Eriksson M, Czene K, Conant EF, et al: Use of low-dose tamoxifen to increase mammographic screening sensitivity in premenopausal women. Cancers 13:302, 2021

39. Niell BL, Augusto B, McIntyre M, et al: Distribution of estimated lifetime breast cancer risk among women undergoing screening mammography. Am J Roentgenol 217:48-55, 2021

40. Eriksson M, Destounis S, Czene K, et al: A risk model for digital breast tomosynthesis to predict breast cancer and guide clinical care. Sci Transl Med 14: eabn3971, 2022

41. Boyd NF, Martin LJ, Yaffe MJ, et al: Mammographic density and breast cancer risk: Current understanding and future prospects. Breast Cancer Res 13: 223-312, 2011

42. Huo CW, Chew GL, Britt KL, et al: Mammographic density—A review on the current understanding of its association with breast cancer. Breast Cancer Res Treat 144:479-502, 2014

43. Weinstein SP, Korhonen K, Cirelli C, et al: Abbreviated breast magnetic resonance imaging for supplemental screening of women with dense breasts and average risk. J Clin Oncol 38:3874-3882, 2020

44. Screening för bröstcancer: Rekommendation och bedömningsunderlag. Artikelnr: 2014-2-32, Stockholm, Sweden, Socialstyrelsen, 2014