cancers

Article

A Clinical Risk Model for Personalized Screening and Prevention of Breast Cancer

Mikael Eriksson [1,2,]*[ID], Kamila Czene [1][ID], Celine Vachon [3], Emily F. Conant [4] and Per Hall [1,5][ID]

[1] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 65 Stockholm, Sweden
[2] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
[3] Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic College of Medicine, Rochester, MN 55905, USA
[4] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[5] Department of Oncology, Södersjukhuset University Hospital, 118 83 Stockholm, Sweden
[*] Correspondence: mikael.eriksson@ki.se; Tel.: +46-852482300

Simple Summary: We investigated the benefits of adding lifestyle and familial risk factors to a mammographic image-derived short-term AI risk model in a 10-year follow-up study for its potential use in personalized screening and prevention of breast cancer (BC). In a case–cohort study, 8110 women were selected from women aged 40–74 participating in a Swedish mammography screening cohort. The women had no BC diagnosis at enrollment. In all, 1661 incident BCs were developed in the case–cohort. The lifestyle/familial-expanded AI risk model showed a significantly higher discriminatory performance in the long term and short term than the imaging-only risk model and the clinical Tyrer–Cuzick v8 model. The expanded model also showed the highest risk classification performance using positive predictive value (PPV). The results suggest that a lifestyle/familial-expanded image-derived AI risk model could most efficiently refine the identification of women who may benefit from personalized screening and/or risk-reducing intervention.

Abstract: Background: Image-derived artificial intelligence (AI) risk models have shown promise in identifying high-risk women in the short term. The long-term performance of image-derived risk models expanded with clinical factors has not been investigated. Methods: We performed a case–cohort study of 8110 women aged 40–74 randomly selected from a Swedish mammography screening cohort initiated in 2010 together with 1661 incident BCs diagnosed before January 2022. The imaging-only AI risk model extracted mammographic features and age at screening. Additional lifestyle/familial risk factors were incorporated into the lifestyle/familial-expanded AI model. Absolute risks were calculated using the two models and the clinical Tyrer–Cuzick v8 model. Age-adjusted model performances were compared across the 10-year follow-up. Results: The AUCs of the lifestyle/familial-expanded AI risk model ranged from 0.75 (95%CI: 0.70–0.80) to 0.68 (95%CI: 0.66–0.69) 1–10 years after study entry. Corresponding AUCs were 0.72 (95%CI: 0.66–0.78) to 0.65 (95%CI: 0.63–0.66) for the imaging-only model and 0.62 (95%CI: 0.55–0.68) to 0.60 (95%CI: 0.58–0.61) for Tyrer–Cuzick v8. The increased performances were observed in multiple risk subgroups and cancer subtypes. Among the 5% of women at highest risk, the PPV was 5.8% using the lifestyle/familial-expanded model compared with 5.3% using the imaging-only model, $p < 0.01$, and 4.6% for Tyrer–Cuzick, $p < 0.01$. Conclusions: The lifestyle/familial-expanded AI risk model showed higher performance for both long-term and short-term risk assessment compared with imaging-only and Tyrer–Cuzick models.

Keywords: breast cancer; risk model; long-term risk; primary prevention; individualized screening; artificial intelligence; image-derived risk model

1. Introduction

A key step in efficient personalized breast cancer screening and risk-reducing interventions is improving breast cancer risk assessment [1,2]. Recent studies have reported that artificial intelligence (AI)-based models using mammograms perform better than traditional lifestyle familial-based risk models in estimating the short-term risk of breast cancer [3–5]. The long-term performance of image-derived models that incorporate lifestyle and family history risk factors has yet to be determined.

Risk assessment for breast cancer is currently offered in the U.S. using lifestyle/familial-based risk tools such as the Tyrer–Cuzick v8 and Gail models [6–8]. Women at a high risk of breast cancer are recommended to undergo supplemental screening with modalities such as breast MRI to increase the early detection of breast cancer. In addition, women at high risk of developing breast cancer may benefit from risk reduction interventions to decrease their risk of developing breast cancer [9–11].

Primary prevention efforts are important because approximately 13% of women are currently diagnosed with breast cancer over their lifetime [12]. Lifestyle changes and medical interventions have shown promising results for reducing breast cancer incidence in high-risk women [13]. Clinical guidelines for high-risk women are also available such as the National Institute for Health and Care Excellence (NICE) and the US Preventive Services Task Force (USPSTF) [14,15].

We used the KARolinska MAmmography Project for Risk Prediction of Breast Cancer (KARMA) screening cohort [16] to investigate the long-term predictive performance using an image-derived AI-based risk model that was expanded with lifestyle risk factors and family history of breast cancer. In this long-term evaluation study, we used an independent set of baseline mammograms from the same underlying screening cohort that was used to develop the model [17]. The AI risk models are available for clinical use in the U.S. and Europe [18].

We studied the predictive performances up to 10 years after study entry to identify women who may benefit from risk-reducing intervention using the lifestyle/familial-expanded model compared to the imaging-only AI risk model, which in turn was compared to the clinically used Tyrer–Cuzick v8 risk model, which includes mammographic density. We also investigated the three models for short-term risk to identify women who may benefit from supplemental screening or a shorter screening interval. We performed analyses in the overall study population and stratified by the risk factors included in the Tyrer–Cuzick v8 risk model to investigate the generalizability of the models to populations with differential risk factor distributions as well as to investigate subgroups of women who could benefit from using the Tyrer–Cuzick v8 model. In addition, we compared the risk models for their ability to capture women who will develop subtype-specific breast cancers.

2. Materials and Methods

2.1. Study Population

In the Swedish national mammography screening program, women aged 40–74 years are invited every 18 or 24 months, depending on age and region [19]. Women who underwent mammographic screening at four hospitals between October 2010 and March 2013 were invited to participate in the prospective KARMA study [16]. Seventy thousand women consented to participate in research on the risk of breast cancer and responded to a web survey on lifestyle and familial-related risk factors for breast cancer. In addition, women donated blood, approved linkages to national registers, and allowed storage and image analysis of their mammograms. In the present study, women with no diagnosis of breast cancer at study entry were eligible. A case–cohort study was conducted consisting of a random subcohort of 8110 (12.1%) from the 66,814 eligible KARMA women together with all incident breast cancers with available risk information that were diagnosed before the register linkage, where 197 of the incident breast cancers were in the subcohort. The article followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines for cohort studies.

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

2.2. Risk Factors at Enrollment and Breast Cancers at Follow-Up

Full-field digital mammographic (FFDM) images were obtained from the left and right breasts (mediolateral oblique (MLO) and cranio-caudal (CC) views) and used to extract AI-based mammographic features (density, microcalcifications, masses, and left–right breast asymmetries of these features) using the ProFound AI Risk (iCAD, Nashua, NH) imaging-only risk tool as previously described (Model 1) [17,20].

The lifestyle/familial-expanded AI risk model used Body Mass Index (BMI), at least monthly use of alcohol, regular smoking, current use of hormone replacement therapy (HRT), and first degree family history of breast cancer in addition to the image-derived features and age as previously described [17]. The factors were extracted from the questionnaire developed as part of the KARMA project [16]. The lifestyle/familial-expanded AI risk model was constructed using the mammographic features in the imaging-only risk model in combination with the lifestyle/familial factors as previously described (Model 2) [17].

The Tyrer–Cuzick v8 risk model was used as a clinical comparison tool [21]. Tyrer–Cuzick v8 provides breast cancer risk based on lifestyle and familial risk factors, i.e., age, height, weight, age at menarche, age at first childbirth, menopausal status, use of hormone replacement therapy (HRT), previous benign breast disorders, first-, second-, and third-degree family history of breast cancer, first- and second-degree family history of ovarian breast cancer, and mammographic density [20]. In our analysis, we did not have access to information regarding history of ovarian cancer in second-degree relatives. We classified mammographic density using the fully automated mammographic density tool STRATUS [20]. Mammographic density was classified into four cBIRADS categories to mimic the Breast Imaging Reporting and Data System (BI-RADS) classification as previously described [22].

Absolute risks of breast cancer were calculated for the three models based on their inclusion of risk factors, risk factor prevalence, and Swedish national statistics on breast cancer incidence rate and competing mortality risk [23]. Risk assessment was performed at study enrollment.

Breast cancer status and mode of detection were retrieved for breast cancers diagnosed up to January 2022 from the National Quality register for Breast Cancer (NKBC) register through linkage using the unique Swedish personal identification numbers [24]. Symptomatic cancer was defined by mode of detection as non-screen-detected cancer. Tumor characteristics were defined using the American Joint Committee on Cancer (AJCC) classification [25].

2.3. Statistical Analysis

Descriptive statistics reported study participant characteristics at study entry [26]. The frequency distribution reported the time from date of mammogram at enrollment to date of breast cancer diagnosis. Absolute risks were estimated at study entry using the three risk models. Inverse probability weights were used to account for the case–cohort sampling [26,27]. Area under the receiver operating characteristics curve (AUC) estimated the discriminatory performance of the models across the 1–10-year follow-up period after age adjustment [28,29]. To estimate the discriminatory performance for a certain year of follow-up, we used the subcohort and the incident breast cancers that developed after study entry between 3 months and that year. The 95% confidence intervals of the AUC point estimates were estimated using 1000 bootstraps [29]. In a subgroup analysis, we also excluded breast cancers diagnosed in the first two years after study baseline. Differences in performances between the lifestyle/familial-expanded risk model and the imaging-only risk model, and in turn between the imaging-only risk model and Tyrer–Cuzick v8, were tested using bootstrapping [30]. The discriminatory performances were reported first for the overall study population and then for subgroups of women stratified by the risk factors included in Tyrer–Cuzick and by breast cancer subtypes (mode of detection, invasiveness,

stage, and receptor status). The lifestyle/familial-expanded AI risk model and Tyrer–Cuzick imputed risk scores for missing risk factor data using built-in methods [17,21]. Manhattan plots and bubble plots presented the significance of risk model AUC differences in the subgroup analyses after adjusting for multiple testing using the Holm–Bonferroni method [31–33]. Positive Predictive Values (PPVs) were estimated with 95% confidence intervals [34]. The proportions of women at low, general, moderate, and high risk based on two-year and ten-year risks were reported according to the National Institute for Health and Care Excellence (NICE) and the U.S. Preventive Service Task Force (USPSTF) guidelines.

Statistical analyses were performed using R 4.1 [35]. All tests were two-sided at significance level 0.05.

3. Results

3.1. Study Population

The case–cohort of 9574 women consisted of 8110 women in the subcohort and 1661 incident breast cancers that were identified in the January 2022 register update as being diagnosed more than three months after enrollment in 2011–2013. A total of 197 out of the 1661 incident breast cancers were found in the subcohort, as shown in Table 1. At study entry, the mean age was 56.52 (±9.53) in cases and 53.86 years (±9.85) in the subcohort. The absolute 2-year risk at baseline was 1.37%, 1.19%, and 0.79% in cases using the lifestyle/familial-expanded AI risk model, the imaging-only AI risk model, and the Tyrer–Cuzick v8 risk model, respectively. The corresponding absolute risks were 0.66%, 0.63%, and 0.60%.

Table 1. Study participant characteristics at enrollment including N = 9574 study participants in the case–cohort with up to 10 years of follow-up.

Characteristic	Subcohort [1], N = 8110 [2]	Cases [1], N = 1661 [2]	*p*-Value [3]
Age	53.88 (9.86)	56.52 (9.53)	<0.01
Height (cm)	166.63 (6.01)	166.93 (5.88)	0.04
Weight (kg)	70.10 (12.54)	70.82 (12.27)	<0.01
BMI	25.24 (4.31)	25.38 (4.26)	0.03
At least monthly alcohol use last year	6461/8054 (80%)	1354/1650 (82%)	0.10
Regular smoking in last year	998/8093 (12%)	197/1658 (12%)	0.65
Age at menarche	13.07 (1.47)	13.11 (1.46)	0.31
Age at first childbirth	27.21 (5.29)	27.20 (5.21)	0.87
Postmenopausal	4411/8110 (54%)	1103/1661 (66%)	<0.01
Current use of HRT [4]	304/7687 (4.0%)	106/1566 (6.8%)	<0.01
Benign breast disease	1756/7930 (22%)	517/1622 (32%)	<0.01
Breast cancer in 1st-degree family			<0.01
No family history	7069/8057 (88%)	1288/1643 (78%)	
Onset age < 50	253/8057 (3.1%)	90/1643 (5.5%)	
Onset age ≥ 50	735/8057 (9.1%)	265/1643 (16%)	
Breast cancer in 2nd/3rd-degree relative	730/5492 (13%)	168/1075 (16%)	0.03
Family history of ovarian cancer	308/7870 (3.9%)	69/1591 (4.3%)	0.36
Percent mammographic density	25.54 (19.45)	28.53 (19.43)	<0.01
Mammographic density above median	3989/8110 (49%)	923/1661 (56%)	<0.01
Imaging-only risk model, 2-year risk score	0.63 (0.95)	1.19 (1.85)	<0.01
Expanded AI risk model, 2-year risk score [5]	0.66 (1.07)	1.37 (2.04)	<0.01
Tyrer–Cuzick v8, 2-year risk score	0.60 (0.39)	0.79 (0.53)	<0.01
Tyrer–Cuzick v8, 10-year risk score	3.09 (1.86)	3.96 (2.49)	<0.01

[1] Subcohort of non-cases and the 197 incident breast cancers diagnosed in the subcohort. For cases, all incident breast cases outside and within the subcohort are included. [2] Mean (SD); n/N (%). [3] Welch Two Sample *t*-test; Pearson's Chi-squared test; Fisher's exact test; Wilcoxon rank sum test. [4] Hormone replacement therapy. [5] Lifestyle/familial-expanded AI risk model.

3.2. Follow-Up of Breast Cancers from Time of Enrollment

The follow-up period for the women in this study was over 10 years and the time from the mammogram at enrollment to breast cancer diagnosis ranged from 3 months to over 10 years, as shown in Figure A1. Half of the cancer events occurred in the first five years. After 10 years of follow-up, 60% of the breast cancers were screen-detected, 31% were diagnosed with a stage 2 or later cancer, 86% were invasive, 86% were estrogen-receptor (ER)-positive, 72% were progesterone-receptor (PR)-positive, and 14% were human epidermal growth factor receptor 2 (HER2)-positive, as shown in Table A1.

3.3. Discriminatory Performance Overall and in Subgroups

The age adjusted discriminatory performance ranged from 0.75 (95%CI 0.70–0.80) after 1 year of follow-up to 0.68 (95%CI 0.66–0.69) after 10 years for the lifestyle/familial-expanded AI risk model, as shown in Figure 1, Table A2. The corresponding 1- and 10-year AUCs were 0.72 (95%CI 0.66–0.78) and 0.65 (95%CI 0.63–0.66) for the imaging-only AI risk model and 0.62 (95%CI 0.55–0.68) and 0.60 (95%CI 0.58–0.61) for Tyrer–Cuzick v8.
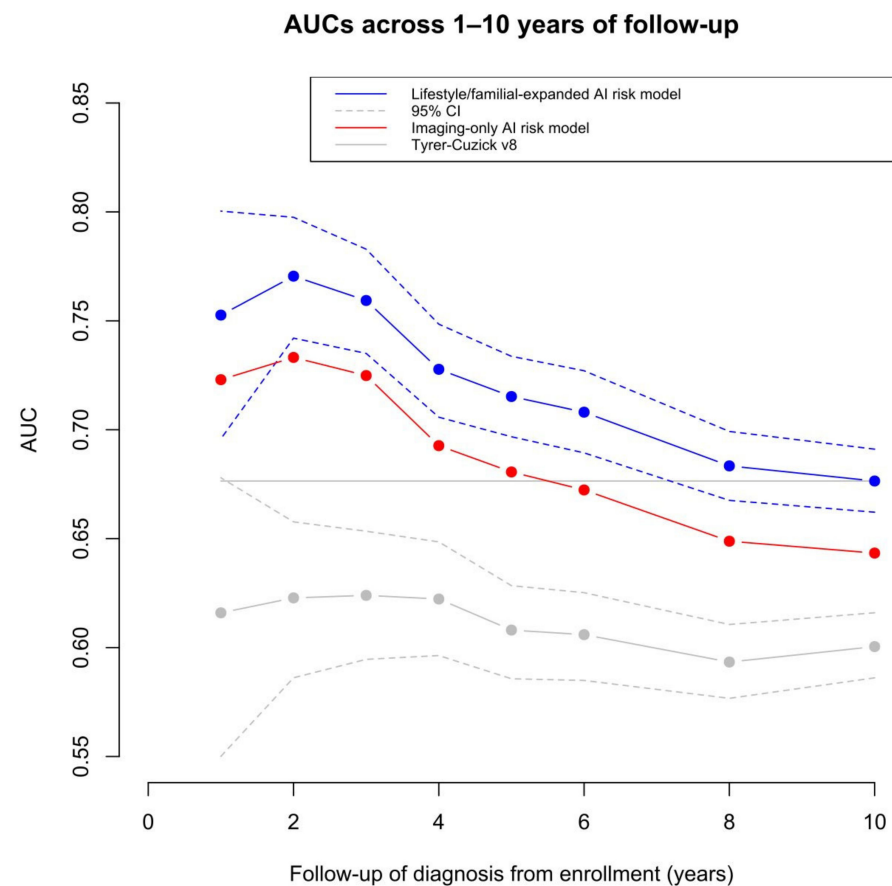

Figure 1. Age adjusted discriminatory performance of the imaging-only AI risk model, the lifestyle/familial-expanded AI risk model, and the Tyrer–Cuzick v8 risk model throughout the first years after enrollment. The 95% confidence intervals are presented for the lifestyle/familial-expanded AI risk model and Tyrer–Cuzick v8 model.

The AUC point estimates for the lifestyle/familial-expanded AI risk model ranged from 0.75 (95%CI 0.70–0.80) after 1 year of follow-up to 0.68 (95%CI 0.66–0.69) after 10 years of follow-up. The corresponding estimates for the imaging-only AI risk model were 0.72 (95%CI 0.66–0.78) and 0.65 (95%CI 0.63–0.66), and for the Tyrer–Cuzick v8 model, they were 0.62 (95%CI 0.55–0.68) and 0.60 (95%CI 0.58–0.61). The horizontal line denotes the AUC point estimate (0.68) of the lifestyle/familial-expanded AI risk model for 10 years of follow-up. When excluding breast cancers diagnosed in the first two years after study baseline, we

observed AUC point estimates for the lifestyle/familial-expanded AI risk model ranging from 0.72 (95%CI 0.67–0.77) after 3 years of follow-up to 0.65 (95%CI 0.61–0.69) after 10 years of follow-up, as shown in Figure A2. The corresponding estimates for the imaging-only AI risk model were 0.70 (95%CI 0.65–0.75) and 0.61 (95%CI 0.57–0.65), and for the Tyrer–Cuzick v8 model, they were 0.63 (95%CI 0.54–0.69) and 0.60 (95%CI 0.58–0.62).

Discriminatory performance tests (N = 160) were performed on risk factor subgroups of women comparing the lifestyle/familial-expanded AI risk model and Tyrer–Cuzick v8 with the imaging-only AI risk model, as shown in Figure 2. The number of subgroups with a significantly different AUC and the strength of significance increased with the number of years of follow-up using the lifestyle/familial-expanded AI risk model, while the corresponding AUCs using Tyrer–Cuzick v8 were similar across the 1–10-year follow-up (Figure 2). Using the lifestyle/familial-expanded AI risk model, all 86 comparisons that were significantly different had a higher AUC, whereas using the Tyrer–Cuzick v8 risk model, all 52 comparisons with a significant difference AUC were lower (Figure A3).

The lifestyle/familial-expanded AI risk model showed significantly higher AUCs compared to the imaging-only AI risk model after 2 years of follow-up, when women with at least monthly intake of alcohol, postmenopausal women, women above median age, below median length, below median weight, below median BMI, above median age at menarche, and above median age at first childbirth benefitted from being assessed with the lifestyle/familial-expanded model compared to the imaging-only AI risk model (Figure A3 bubble plot). After 10 years of follow-up, all nine risk factors (alcohol intake, menopause, age, height, weight, BMI, age at menarche, age at first childbirth, benign breast disease) were significant. Women with a family history of breast cancer did not have a significantly increased AUC over the imaging-only AI risk model, nor did women with second/third-degree breast cancer family history or ovarian cancer in the family (Figure A3).

(A). Lifestyle/familial-expanded AI risk model vs. imaging-only AI risk model (160 comparisons)
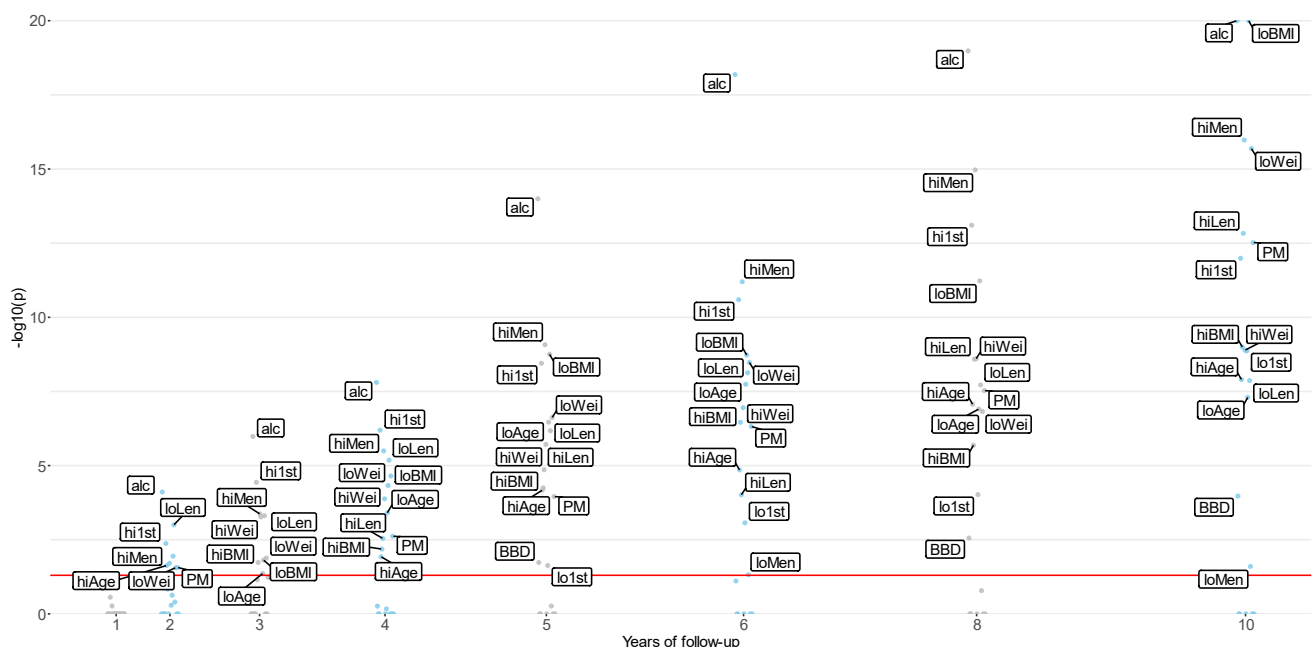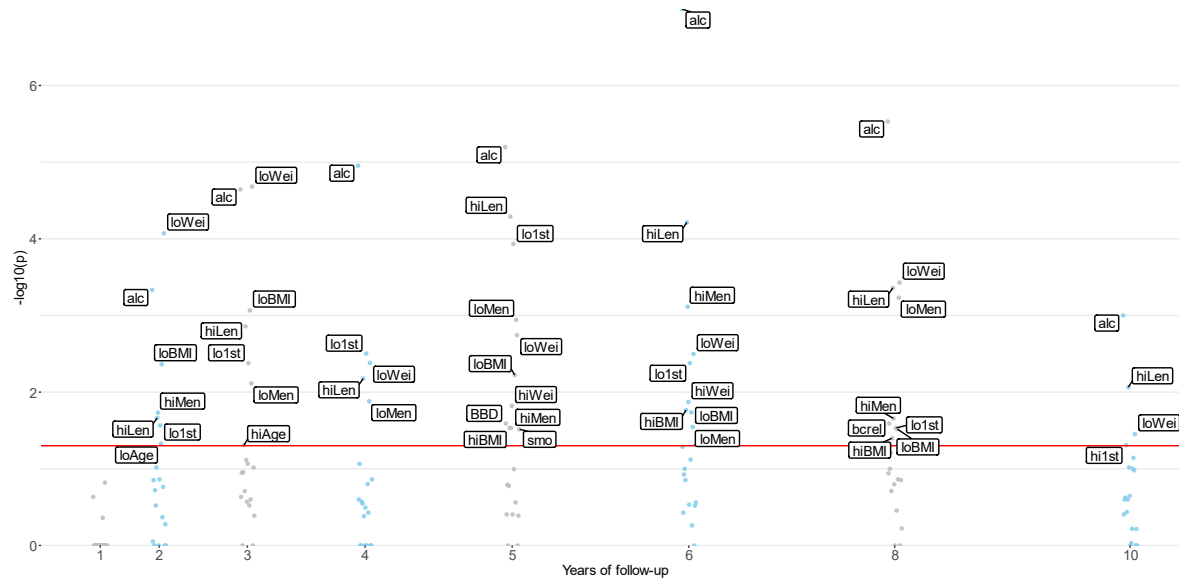

Figure 2. *Cont.*

(B). Tyrer–Cuzick v8 model vs. imaging-only AI risk model (160 comparisons)

Legend

alc—at least monthly use of alcohol

smo—currently smoking

hi/loAge—above/below median age

hi/loLen—above/below median length

hi/loWei—above/below median weight

hi/loBMI—above/below median BMI

hi/loMen—above/below median age at menarche

AUC—Area Under the receiver operating characteristics Curve

hi/lo1st—above/below median age at first childbirth

PM—postmenopausal

hrt—currently using hormone replacement therapy

BBD—benign breast disease

FH—1st degree of family history

bcr—2nd/3rd degree of family history

OV—ovarian cancer in family

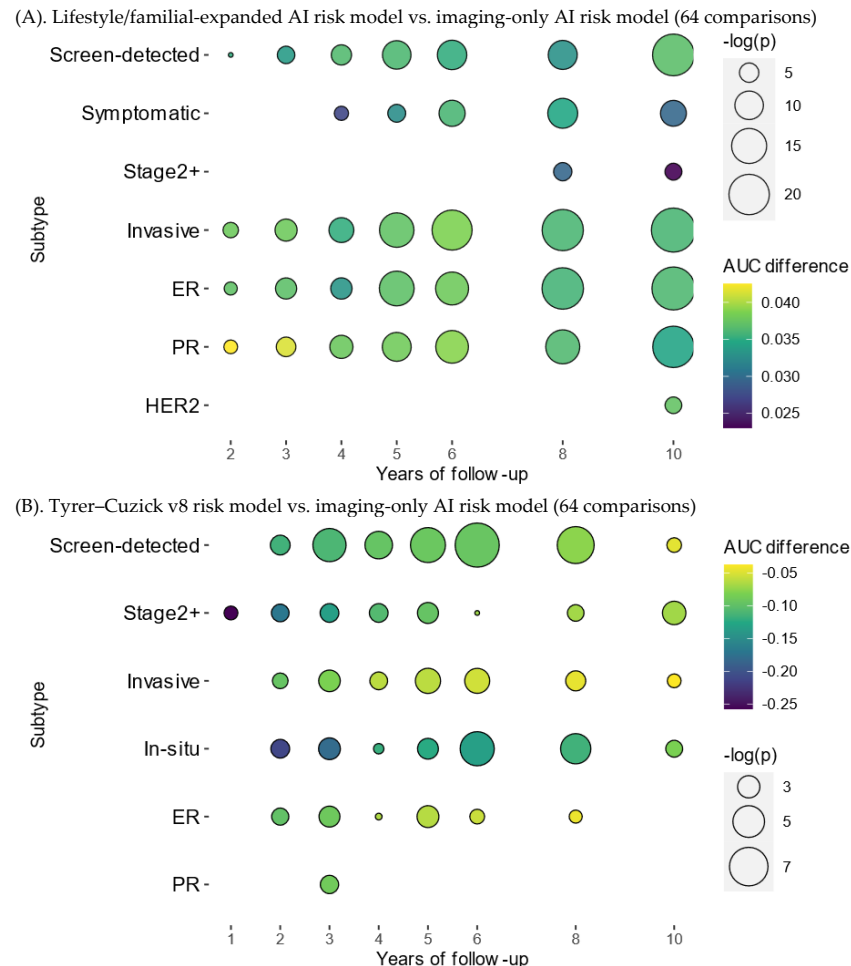hi/loPD—above/below median percent density

Figure 2. Manhattan plot of *p*-values denoting the significance of age adjusted AUC differences across 1–10 years of follow-up in subgroups of women by risk factor when comparing the lifestyle/familial-expanded AI risk model with the imaging-only AI risk model, which in turn is compared with Tyrer–Cuzick v8. The subgroups of women by risk factor are above/below median: age, length, weight, Body Mass Index (BMI); at least monthly intake of alcohol, current regular smoker; above/below median age at menarche, age at first childbirth, mammographic density; menopause, current use of hormone replacement therapy (HRT), benign breast disease, family history of breast cancer, 2nd/3rd degree relative with breast cancer, and ovarian cancer in the family. The red horizontal line represents the significance threshold (*p* = 0.05) after adjusting the *p*-values for multiple comparison using the Holm–Bonferroni method.

3.4. Predictive Value by Tumor Characteristics

The age adjusted discriminatory performance estimations by breast cancer subtypes resulted in 64 tests comparing the lifestyle/familial-expanded AI risk model with the imaging-only AI risk model, which in turn was compared with Tyrer–Cuzick v8 (Figure A4). Similar tendencies were observed with increasing number of significant AUC differences and increasing years of follow-up for the lifestyle/familial-expanded model, while AUC differences were not depending on time for Tyrer–Cuzick v8. All 36 significant comparisons for the lifestyle/familial-expanded model showed higher AUCs, in contrast to all 36 significant comparisons for the Tyrer–Cuzick v8 risk model, which showed lower AUCs (Figure A5).

The significant discriminatory improvements of the lifestyle/familial-expanded AI risk model over the imaging-only AI risk model ranged from 2.3–4.3 percentage points for screen-detected, symptomatic, invasive, ER-, PR-, HER2-positive, and stage 2 or later

breast cancers (Figure 3, Table A3). Screen-detected, invasive, ER-, and PR-positive breast cancers were significant from 2 years of follow-up, while stage 2 or later cancers and HER2 cancers were significant after 8–10 years of follow-up. For Tyrer–Cuzick v8, screen-detected, invasive, in situ, ER- and PR-positive, and stage 2 or later breast cancers had significantly lower AUCs (3.7–25.8 percentage points) compared with the imaging-only model at one or several years in the 1–10-year follow-up (Figure 3, Table A4).
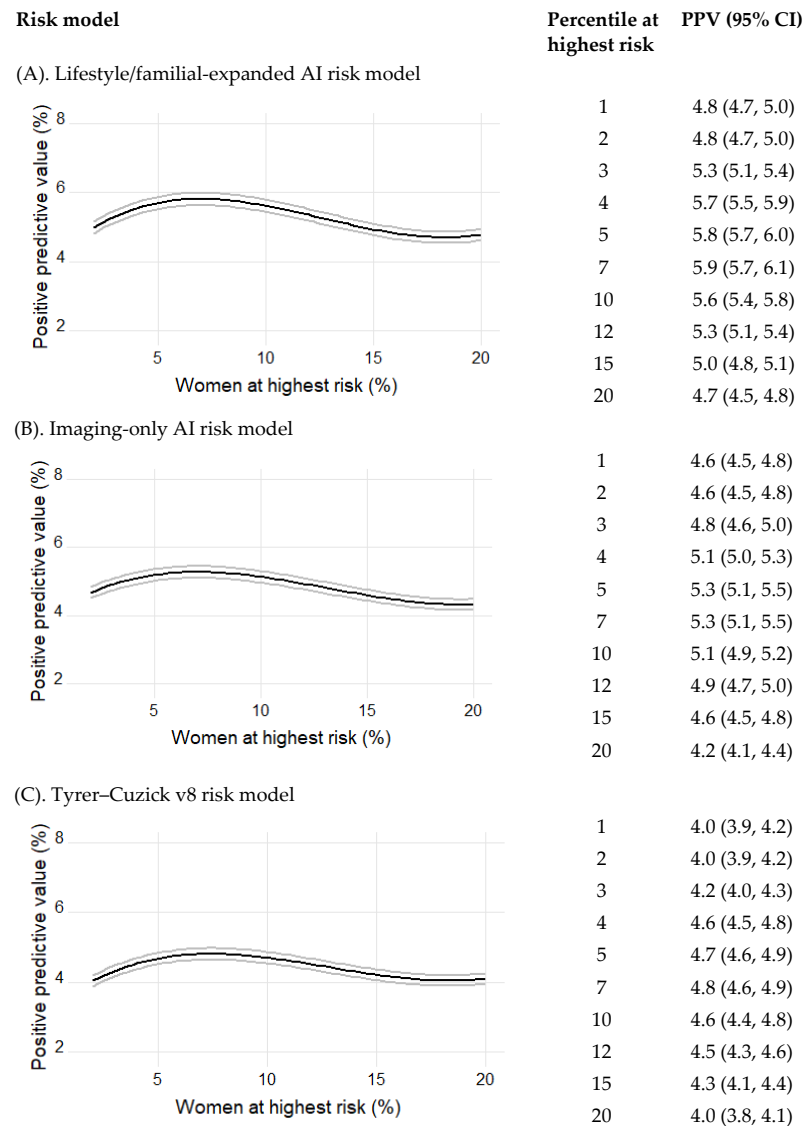
(A). Lifestyle/familial-expanded AI risk model vs. imaging-only AI risk model (64 comparisons)

(B). Tyrer–Cuzick v8 risk model vs. imaging-only AI risk model (64 comparisons)

AUC—Area Under the receiver operating characteristics Curve. ER—ER-positive. PR—PR-positive. HER2—HER2-positive. Stage2+—stage 2 or later

Figure 3. Bubble plot of *p*-values and AUC differences across 1–10 years of follow-up in subtypes of cancers comparing the lifestyle/familial-expanded AI risk model with the imaging-only AI risk model, which in turn is compared with Tyrer–Cuzick v8. The plot presents significantly different AUCs ($p < 0.05$) that were extracted after Holm–Bonferroni adjustment for multiple comparison. The bubble sizes represent minus-log-transformed *p*-values and the bubble colors represent age adjusted AUC point estimate differences. The investigated subtypes are mode of detection (screen-detected, symptomatic), stage (stage 1 or earlier, stage 2 or later), invasiveness (invasive, in situ), estrogen-receptor (ER), progesterone-receptor (PR), and human epidermal growth factor receptor 2 (HER2).

3.5. Positive Predictive Value (PPV) and Clinical Risk Classification

PPVs were estimated in multiple subgroups of women ranging from 1% to 20% of women at highest risk of breast cancer for each of the three models. Using the lifestyle/familial-expanded AI risk model, the PPVs ranged from 4.8% (95%CI 4.7–5.0) to 4.7% (95%CI 4.5–4.8) across the 1–20% of women at the highest risk (Figure 4). The corresponding PPVs for the imaging-only AI risk model were 4.6% (95%CI 4.5–4.8) and 4.2% (95%CI 4.1–4.4) and for Tyrer–Cuzick v8, they were 4.0% (95%CI 3.9–4.2) and

4.0% (95%CI 3.8–4.1). The 5% of women at highest risk had a PPV of 5.8% for the lifestyle/familial-expanded AI risk model compared with 5.3% for the imaging-only AI risk model, $p < 0.01$, and 4.6% for Tyrer–Cuzick v8, $p < 0.01$.

Risk model	Percentile at highest risk	PPV (95% CI)
(A). Lifestyle/familial-expanded AI risk model		
	1	4.8 (4.7, 5.0)
	2	4.8 (4.7, 5.0)
	3	5.3 (5.1, 5.4)
	4	5.7 (5.5, 5.9)
	5	5.8 (5.7, 6.0)
	7	5.9 (5.7, 6.1)
	10	5.6 (5.4, 5.8)
	12	5.3 (5.1, 5.4)
	15	5.0 (4.8, 5.1)
	20	4.7 (4.5, 4.8)
(B). Imaging-only AI risk model		
	1	4.6 (4.5, 4.8)
	2	4.6 (4.5, 4.8)
	3	4.8 (4.6, 5.0)
	4	5.1 (5.0, 5.3)
	5	5.3 (5.1, 5.5)
	7	5.3 (5.1, 5.5)
	10	5.1 (4.9, 5.2)
	12	4.9 (4.7, 5.0)
	15	4.6 (4.5, 4.8)
	20	4.2 (4.1, 4.4)
(C). Tyrer–Cuzick v8 risk model		
	1	4.0 (3.9, 4.2)
	2	4.0 (3.9, 4.2)
	3	4.2 (4.0, 4.3)
	4	4.6 (4.5, 4.8)
	5	4.7 (4.6, 4.9)
	7	4.8 (4.6, 4.9)
	10	4.6 (4.4, 4.8)
	12	4.5 (4.3, 4.6)
	15	4.3 (4.1, 4.4)
	20	4.0 (3.8, 4.1)

PPV—Positive predictive value

Figure 4. Positive predictive values of women at the highest risk of breast cancer after 10 years of follow-up. The plots present the positive predictive value (the proportion of breast cancers) captured by the model at study baseline among the 1–20% of women at the highest risk of breast cancer as predicted by the image-based risk model and the Tyrer–Cuzick v8 risk model, respectively.

The proportion of breast cancers that were identified as high-risk per NICE guidelines using the lifestyle/familial-expanded AI risk model (22%) was significantly larger than the corresponding proportions identified using the imaging-only AI risk model (19%), $p < 0.01$, and Tyrer–Cuzick v8 (7.2%) with 10-year risk projection, $p < 0.01$ (Table A5). The high-risk women were 6.7 times more likely to be diagnosed with breast cancer compared to women at general risk using the lifestyle/familial-expanded model (Table A5). The corresponding figure for the imaging-only model was 6.4 and, for Tyrer–Cuzick v8, 4.2. Similar differences in risk stratification performances were observed using USPSTF guidelines (Table A6).

The positive predictive value is presented after inverse probability weighting to account for the case–cohort sampling with 95% confidence intervals.

Tests were performed to determine the difference between the binomial proportions of breast cancers captured at high risk by the lifestyle/familial-expanded AI risk model, the imaging-only AI risk model, and Tyrer–Cuzick v8. In the 5% of women at the highest risk, the proportion of cancer that was captured by the lifestyle/familial-expanded AI risk model (5.8%) was significantly higher than the 5.3% captured by the image-based risk model, $p < 0.01$, which in turn was significantly higher than the 4.6% of breast cancers that were captured by Tyrer–Cuzick v8 10-year, $p < 0.01$. The corresponding numbers for the 12% of women at highest risk were 5.3 and 4.9, $p < 0.01$, and 4.5, $p < 0.01$.

4. Discussion

We investigated the performance of a lifestyle/familial-expanded AI risk model over an imaging-only-derived AI risk model, in turn, over the clinical Tyrer–Cuzick v8 risk model within a large-scale prospective screening cohort. The lifestyle/familial-expanded AI risk model showed significantly better discriminatory performances of 2–4% over the imaging-only AI risk model, which in turn performed 4–26% better than Tyrer–Cuzick v8, across multiple subgroups of women defined by breast cancer risk factors and by cancer subtypes.

The increase in discriminatory performance using the lifestyle/familial-expanded AI risk model over the imaging-only AI risk model could improve personalized screening and risk-reducing interventions. At the same time, an imaging-only risk algorithm provides several advantages over a model that requires data from additional sources. Image data are available for all women attending screening and an image-based algorithm provides consistent scores across a screening population. A lifestyle/familial-expanded AI risk model may add additional time and costs for implementation and is based on self-reporting prone to missing data and recall bias.

We estimated the model performance differences in subgroups using dichotomized risk factors to enable higher sample sizes and to decrease the likelihood of reporting false positive/negative results [36]. The lifestyle/familial-expanded AI risk model predicted risks non-significantly different compared to the imaging-only AI risk model in most of the investigated risk subgroups. This indicates that women with increased lifestyle/familial risk exposures also have increased exposures of the mammographic features in the imaging-only AI risk model. However, several subgroups of women defined by risk factors (alcohol intake, menopause status, age at menarche, height, BMI, age at first childbirth, and benign breast disease) were better identified using the lifestyle/familial-expanded AI risk model.

Compared with the imaging-only AI risk model, the discriminatory performance of the Tyrer–Cuzick v8 model was not significantly different in approximately three quarters of all investigated subgroups. In the remaining subgroups, the AUCs were significantly lower in the range of 4–26 percent with the largest AUC difference for stage 2 or later cancers.

In the U.S., approximately 12% of women of screening age have a lifetime risk of ≥20% and are eligible for supplemental screening with breast MRI [37]. In our study, we found that the PPV for risk prediction was significantly lower for Tyrer–Cuzick v8 than for lifestyle/familial-expanded image-derived AI risk model in 12% of women at the highest risk. We did not find any subgroup of women who had a higher discriminatory performance using the Tyrer–Cuzick v8 compared to the imaging-only model or the lifestyle/familial-expanded AI risk model. In addition, we did not find that women with a family history of breast cancer had significantly increased discriminatory performance using the lifestyle/familial-expanded AI risk model compared to the imaging-only AI risk model.

The lifestyle/familial-expanded AI risk model showed an advantage over the imaging-only AI risk model for identifying ER-positive breast cancers across the 1–10-year follow-up. Women therefore could have an advantage in having their risk assessed using the lifestyle/familial-expanded model when considering prophylactic endocrine therapy to reduce their risk of developing breast cancer [38].

Image-based models have been reported to perform well in the short term and are designed to personalize breast cancer screening [3,17,39]. Risk assessment in the short

term identifies women who may benefit from supplemental screening or from a shorter screening interval to help identify breast cancers at an earlier stage. We did not see a significant advantage in identifying symptomatic and stage 2 or later cancers in the short term using the lifestyle/familial-expanded model over the imaging-only AI risk model.

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 [14,15]. Considering that we observed a considerable and consistent discriminatory advantage using an image-derived AI-based risk model over a traditional lifestyle/familial-based risk model, investigations should evaluate extending guidelines to incorporate such newer, image-derived risk tools.

This study has several limitations. We investigated the model performances in a Swedish large-scale screening population for women at general risk, where the vast majority of women were white, attended biennial screening using digital mammography (GE, Philips, Sectra, Siemens, Fuji machines), and had a recall rate of ~3% with a cancer detection rate of ~5/1000 exams [40]. Our comparison with Tyrer–Cuzick v8 was limited by the fact that we did not have information on ovarian cancer in second-degree relatives and that we did not use a continuous measure for mammographic density. Further studies are needed investigating the generalizability of our reported results in independent screening programs. Further studies are also needed to investigate the extent of existing versus developing breast cancers in a ten-year time range after baseline. This may have consequences for strategies for improving breast cancer screening versus primary prevention. Studies are also needed to compare the lifestyle/familial-expanded AI risk model with the Tyrer–Cuzick v8 model and to investigate the addition of genetic determinants of breast cancer to the risk model.

5. Conclusions

A lifestyle/familial-expanded image-derived AI-based risk model showed a higher performance for capturing women who may benefit from risk reduction intervention and/or shorter screening intervals or supplemental screening compared to an imaging-only AI-based risk model and the clinical Tyrer–Cuzick v8 model. The higher performance was observed between 1 and 10 years after risk assessment across multiple subgroups of women defined by breast cancer risk factors and by cancer subtypes.

Appendix A

Figure A1. Time from mammogram at enrollment to breast cancer diagnosis. N = 1661 breast cancer cases were assessed for risk of breast cancer at enrollment. Min: 0.3 years, median: 4.8, mean: 5.2, max: 11.2 years.
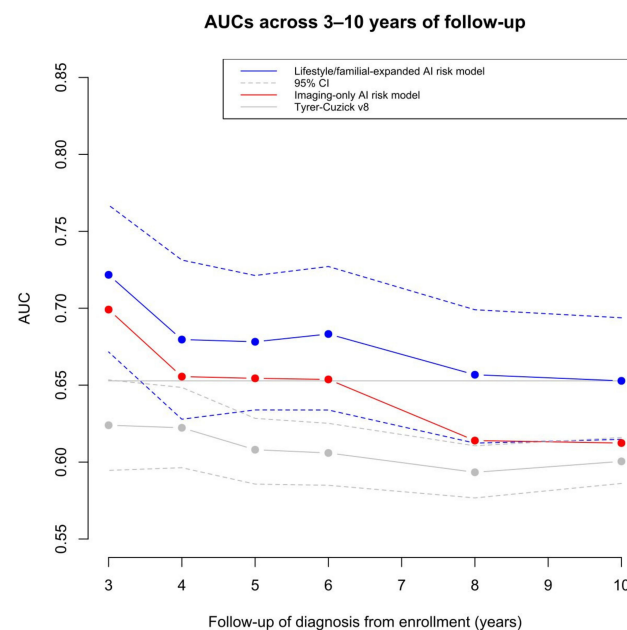
Figure A2. Age adjusted discriminatory performance of the imaging-only AI risk model, the lifestyle/familial-expanded AI risk model, and the Tyrer–Cuzick v8 risk model 3–10 years after enrollment. The 95% confidence intervals are presented for the lifestyle/familial-expanded AI risk model and Tyrer–Cuzick v8 model.

(A). Lifestyle/familial-expanded AI risk model vs. imaging-only AI risk model (160 comparisons)

(B). Tyrer–Cuzick v8 risk model vs. imaging-only AI risk model (160 comparisons)

1. Table of model performance differences

Adjusted *p*-value	Higher AUC	Lower AUC	Adjusted *p*-value	Higher AUC	Lower AUC
<0.05	86	0	<0.05	0	52
0.05–<0.25	7	0	0.05–<0.25	0	38
0.25–<0.5	4	0	0.25–<0.5	0	26
0.5–<0.75	5	0	0.5–<0.75	0	5
≥0.75	52	6	≥0.75	0	39

2. Bubble plot of significant *p*-values across follow-up by risk factor

Legend for risk factors

alc—at least monthly use of alcohol
smo—currently smoking
hi/loAge—above/below median age
hi/loLength—above/below median length
hi/loWeight—above/below median weight
hi/loBMI—above/below median BMI
hi/loMen—above/below median age at menarche
hi/lo1st—above/below median age at first childbirth

PM—postmenopausal
hrt—currently using hormone replacement therapy
BBD—benign breast disease
FH—1st degree of family history
bcr—2nd/3rd degree of family history
OV—ovarian cancer in family
hi/loPD—above/below median percent density

Figure A3. Significance of risk model age adjusted AUC differences by risk factor subgroups. Comparisons of risk model AUC differences across 1–10 years of follow-up and risk factors. The lifestyle/familial-expanded AI risk model and Tyrer–Cuzick v8 were compared with the imaging-only AI risk model, respectively. The examined risk factors were above or below the median: age, length, weight, BMI; at least monthly intake of alcohol, current regular smoker; above/below median age at menarche, age at first childbirth, mammographic density; menopause, current use of HRT, benign breast disease, family history of breast cancer, 2nd/3rd-degree relative with breast cancer, and ovarian cancer in the family. The table presents the number of tests at different *p*-value thresholds after adjusting for multiple testing using the Holm–Bonferroni method. The bubble plot presents the minus-log-transformed *p*-value significance across the follow-up period for each significant risk factor.

When comparing the lifestyle/familial-expanded risk model with the imaging-only risk model, there were no significant AUC differences after any years of follow-up for current regular smokers, above and below median percentage area of mammographic density, premenopausal women, current users of HRT, family history of breast cancer, second/third-degree relative with breast cancer, and ovarian cancer in the family.

When comparing the Tyrer–Cuzick v8 risk model with the imaging-only risk model, there were no significant AUC differences after any years of follow-up for menopause status, above and below median percentage area of mammographic density, premenopausal women, current users of HRT, family history of breast cancer, and ovarian cancer in the family.

(A). Lifestyle/familial-expanded AI risk model vs. imaging-only AI risk model (64 comparisons)
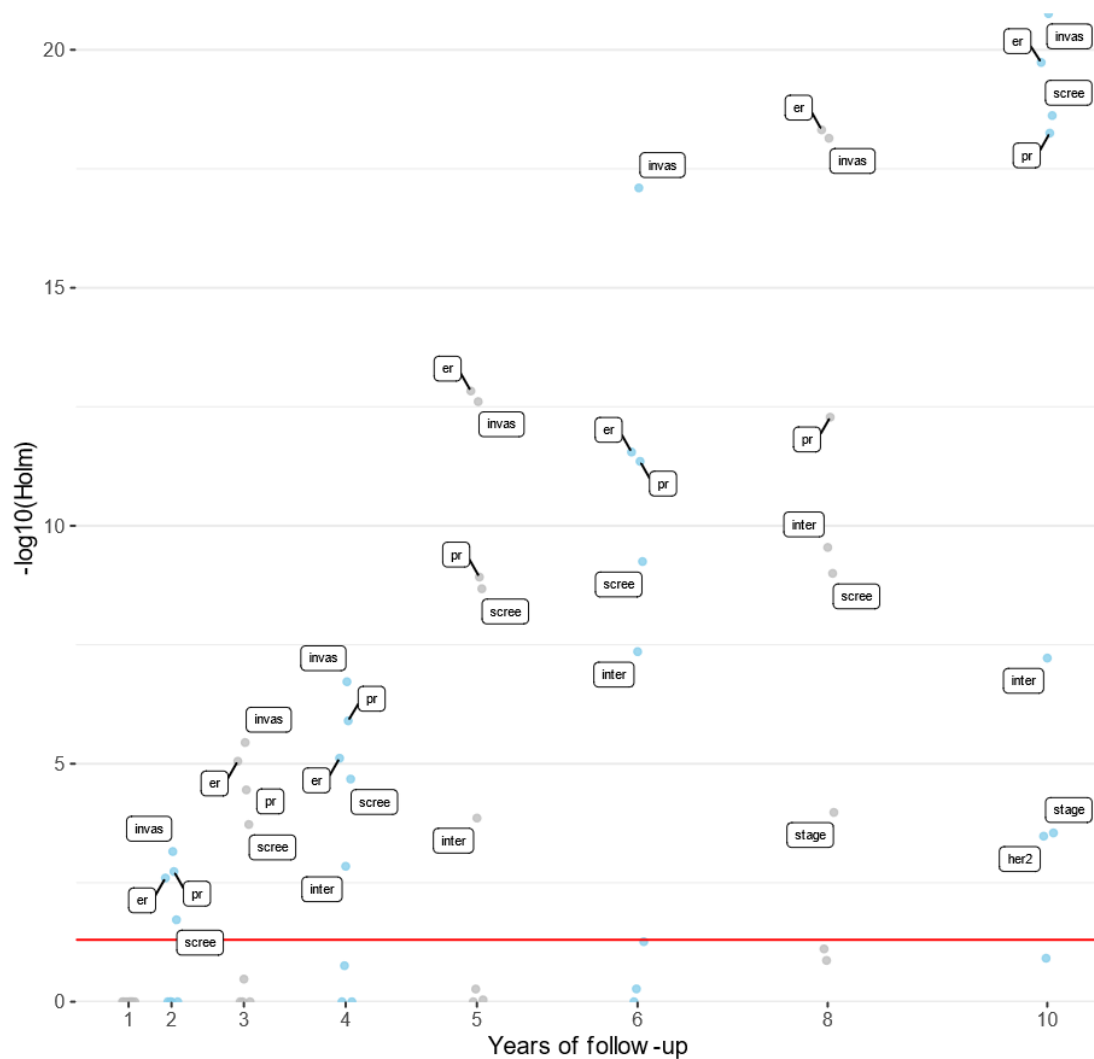
Figure A4. *Cont.*

(B). Tyrer–Cuzick v8 risk model vs. imaging-only AI risk model (64 comparisons)
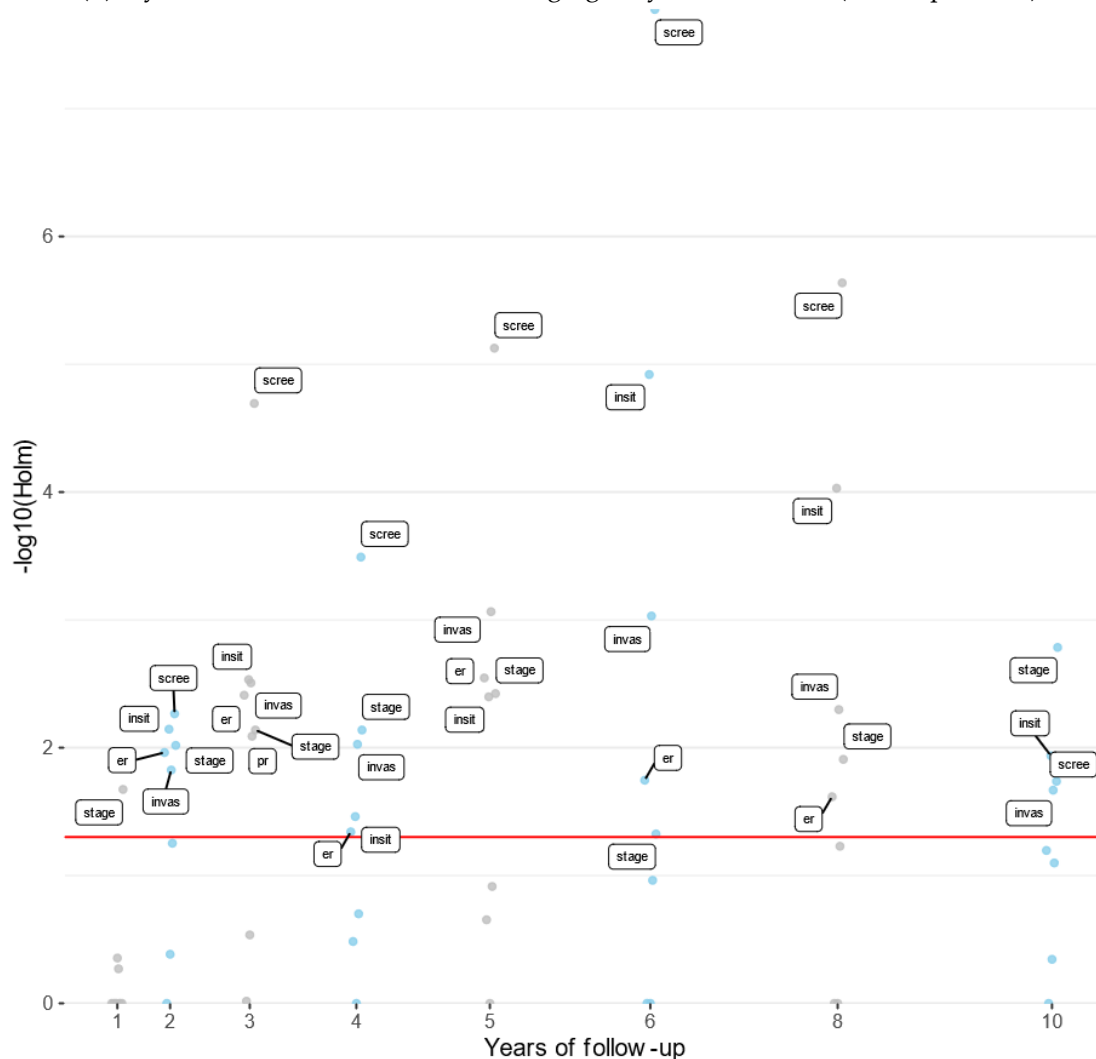
Legend

scree—screen detected breast cancer

inter—symptomatic cancer	er—estrogen-positive
stage—stage 2 or later	pr—progesterone-positive
invasive—invasive breast cancer	her2—HER2-positive
insitu—in situ breast cancer	

Figure A4. Manhattan plot of risk model differences by breast cancer subtypes and years of follow-up. Manhattan plot of the significance of age adjusted AUC differences across 1–10 years of follow-up for subgroups of women defined by risk factors comparing the lifestyle/familial-expanded AI risk model and Tyrer–Cuzick v8 with the imaging-only AI risk model. The plotted labels were limited to AUC comparisons with a *p*-value below 0.01 after adjustment for multiple comparisons using the Holm–Bonferroni method.

(A). Lifestyle/familial-expanded AI risk model vs. imaging-only AI risk model (64 comparisons) (B). Tyrer–Cuzick v8 risk model vs. imaging-only AI risk model (64 comparisons)

1. Table of model performance differences

Adjusted *p*-value	Higher AUC	Lower AUC	Adjusted *p*-value	Higher AUC	Lower AUC
<0.05	36	0	<0.05	0	36
0.05–<0.25	5	1	0.05–<0.25	0	8
0.25–<0.5	1	0	0.25–<0.5	0	5
0.5–<0.75	2	0	0.5–<0.75	0	1
≥0.75	19	1	≥0.75	0	14

2. Bubble plot of significant *p*-values across follow-up by cancer subtype
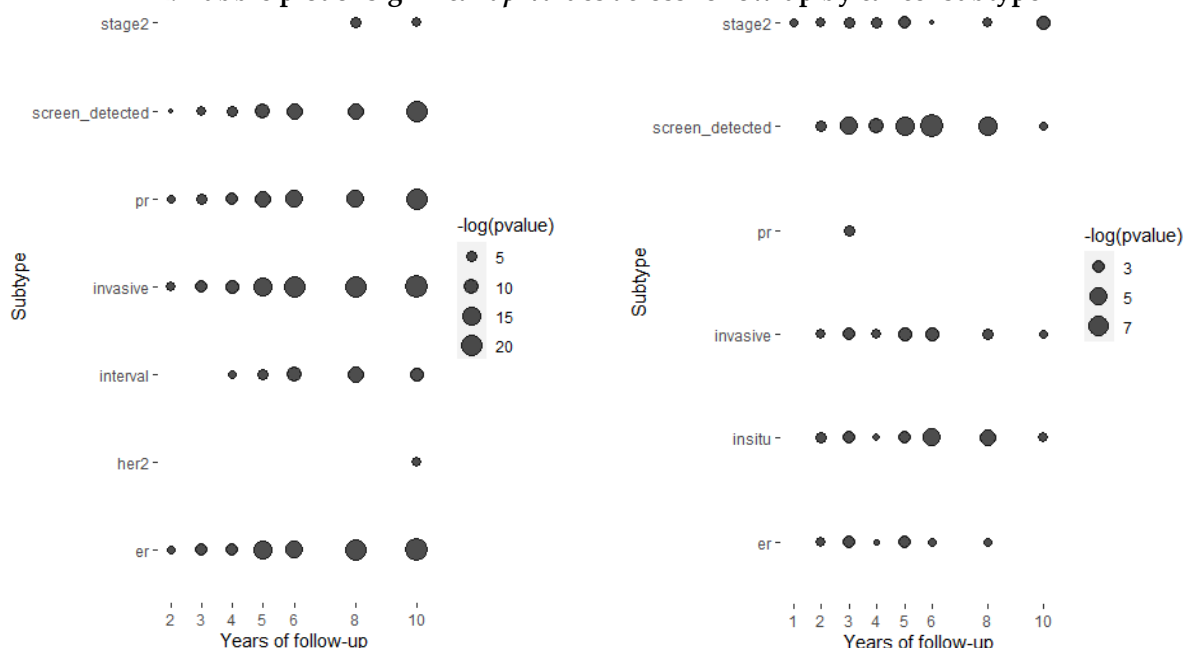
Legend for subtypes

screen—screen detected breast cancer
interval—symptomatic cancer
stage2—stage 2 or later
invasive—invasive breast cancer
insitu—in situ breast cancer

er—estrogen-positive
pr—progesterone-positive
her2—HER2-positive

Figure A5. Significance of risk model age adjusted AUC differences by cancer subtype. Comparisons of risk model AUC differences across 1–10 years of follow-up and cancer subtypes. The lifestyle/familial-expanded AI risk model and Tyrer–Cuzick v8 were compared with the imaging-only AI risk model, respectively. The examined subtypes were mode of detection (screen-detected, symptomatic), stage (stage 1 or earlier, stage 2 or later), invasiveness (invasive, in situ), and receptor status (ER, PR, HER2). The table presents the number of tests at different *p*-values thresholds after adjusting for multiple testing using the Holm–Bonferroni method. The bubble plot presents the minus-log-transformed *p*-value significance across the follow-up period for each significant breast cancer subtype.

When comparing the lifestyle/familial-expanded AI risk model with the imaging-only risk model, there were no significant AUC differences after any years of follow-up for in situ breast cancers.

When comparing the Tyrer–Cuzick v8 risk model with the imaging-only risk model, there were no significant AUC differences after any years of follow-up for symptomatic and HER2 breast cancers.

Appendix B

Table A1. Mode of detection, invasiveness, stage, and receptor status in breast cancer cases. Frequencies and percentages of breast cancers diagnosed during the 10-year follow-up.

Characteristic	n/N (%)
Mode of detection	
Screen detected	978/1635 (60%)
Symptomatic cancers	657/1635 (40%)
(Missing)	26
Tumor invasiveness	
Invasive	1391/1620 (86%)
In situ	229/1620 (14%)
(Missing)	41
Stage	
Stage 2 or later	504/1633 (31%)
Stage 0 or 1	1129/1633 (69%)
(Missing)	28
ER status	
ER-positive	1217/1414 (86%)
ER-negative	197/1414 (14%)
(Missing)	247
PR status	
PR-positive	1019/1412 (72%)
PR-negative	393/1412 (18%)
(Missing)	249
HER2 status	
HER2-positive	200/1391 (14%)
HER2-negative	1191/1391 (86%)
(Missing)	270

ER—estrogen-receptor, PR—progesterone-receptor, HER2—human epidermal growth factor receptor 2.

Table A2. Age adjusted discriminatory performances of three risk models stratified by risk factor subgroups. Discriminatory performance AUC of the lifestyle/familial-expanded AI risk model, the imaging-only AI risk model, and Tyrer–Cuzick v8 in subgroups of women by use of alcohol, smoking, BMI, age at menarche, age at first birth, menopause, use of HRT, benign breast disease, and family histories. Multiple estimations were performed for breast cancers diagnosed 1–10 years after study entry. Risk groups with high and low risk exposures represent above and below median risk exposures, respectively.

Group of Women	1 Year	2 Years	3 Years	4 Years	5 Years	6 Years	8 Years	10 Years
Expanded AI risk model [2]								
All women	0.75	0.77	0.76	0.73	0.72	0.71	0.68	0.68
Use of alcohol [1]	0.76	0.77	0.76	0.73	0.72	0.71	0.69	0.68
Benign breast disease	0.66	0.77	0.76	0.74	0.72	0.71	0.68	0.68
2nd/3rd-degree breast cancer	0.68	0.77	0.77	0.75	0.72	0.70	0.67	0.66
Family history of breast cancer	0.67	0.68	0.68	0.66	0.66	0.65	0.63	0.63
High age at first childbirth	0.72	0.78	0.75	0.73	0.72	0.71	0.68	0.68
High age	0.70	0.75	0.75	0.73	0.71	0.71	0.69	0.69
High BMI	0.73	0.77	0.77	0.75	0.74	0.73	0.70	0.69
High length	0.71	0.79	0.77	0.73	0.71	0.71	0.68	0.68
High menarche	0.74	0.76	0.76	0.72	0.72	0.71	0.69	0.68
High weight	0.74	0.77	0.77	0.74	0.73	0.73	0.70	0.69
Use of HRT	0.51	0.70	0.73	0.72	0.70	0.69	0.65	0.63
Low age at first childbirth	0.76	0.76	0.76	0.72	0.70	0.70	0.67	0.67
Low age	0.76	0.79	0.76	0.72	0.72	0.70	0.66	0.65

Table A2. *Cont.*

Group of Women	1 Year	2 Years	3 Years	4 Years	5 Years	6 Years	8 Years	10 Years
Low BMI	0.77	0.77	0.75	0.70	0.69	0.68	0.67	0.67
Low length	0.76	0.75	0.75	0.73	0.72	0.71	0.68	0.68
Low menarche	0.70	0.77	0.76	0.73	0.71	0.70	0.67	0.67
Low weight	0.74	0.77	0.75	0.71	0.70	0.68	0.66	0.66
Ovarian cancer in family	0.60	0.77	0.73	0.66	0.68	0.67	0.69	0.68
Postmenopausal	0.74	0.75	0.75	0.72	0.71	0.71	0.69	0.68
Current smoker	0.64	0.79	0.72	0.72	0.71	0.69	0.68	0.67
Imaging-only AI risk model								
All women	0.72	0.73	0.73	0.69	0.68	0.67	0.65	0.64
Use of alcohol [1]	0.71	0.73	0.72	0.69	0.68	0.67	0.65	0.64
Benign breast disease	0.68	0.75	0.74	0.72	0.70	0.69	0.66	0.65
2nd/3rd-degree breast cancer	0.70	0.73	0.75	0.73	0.69	0.68	0.67	0.65
Family history of breast cancer	0.64	0.68	0.68	0.64	0.65	0.64	0.62	0.62
High age at first childbirth	0.66	0.73	0.70	0.68	0.67	0.66	0.64	0.64
High age	0.69	0.72	0.72	0.70	0.68	0.68	0.66	0.66
High BMI	0.71	0.74	0.73	0.72	0.71	0.70	0.67	0.66
High length	0.71	0.77	0.75	0.70	0.68	0.68	0.65	0.64
High menarche	0.71	0.73	0.72	0.68	0.68	0.67	0.65	0.64
High weight	0.70	0.73	0.73	0.71	0.70	0.70	0.67	0.66
Use of HRT	0.54	0.73	0.75	0.70	0.69	0.67	0.63	0.61
Low age at first childbirth	0.74	0.73	0.73	0.70	0.67	0.67	0.64	0.63
Low age	0.71	0.74	0.71	0.67	0.67	0.65	0.62	0.62
Low BMI	0.73	0.72	0.72	0.66	0.65	0.65	0.63	0.63
Low length	0.71	0.69	0.70	0.69	0.68	0.67	0.65	0.64
Low menarche	0.67	0.74	0.74	0.72	0.69	0.68	0.66	0.65
Low weight	0.71	0.73	0.72	0.68	0.66	0.64	0.63	0.62
Ovarian cancer in family	0.56	0.71	0.69	0.62	0.64	0.65	0.67	0.67
Postmenopausal	0.72	0.72	0.72	0.70	0.68	0.68	0.66	0.65
Current smoker	0.61	0.75	0.72	0.72	0.71	0.69	0.67	0.66
Tyrer–Cuzick v8								
All women	0.62	0.62	0.62	0.62	0.61	0.61	0.59	0.60
Use of alcohol [1]	0.59	0.61	0.61	0.61	0.60	0.60	0.59	0.60
Benign breast disease	0.61	0.65	0.64	0.63	0.61	0.61	0.59	0.60
2nd/3rd-degree breast cancer	0.63	0.53	0.57	0.57	0.58	0.57	0.55	0.57
Family history of breast cancer	0.52	0.60	0.61	0.58	0.56	0.55	0.55	0.56
High age at first childbirth	0.60	0.64	0.62	0.63	0.62	0.61	0.59	0.59
High age	0.63	0.64	0.63	0.64	0.62	0.63	0.61	0.62
High BMI	0.64	0.65	0.65	0.66	0.64	0.64	0.62	0.62
High length	0.64	0.65	0.63	0.62	0.60	0.60	0.59	0.59
High menarche	0.61	0.63	0.63	0.63	0.61	0.60	0.60	0.60
High weight	0.69	0.66	0.66	0.66	0.64	0.64	0.62	0.62
Use of HRT	0.46	0.48	0.54	0.52	0.55	0.54	0.57	0.58
Low age at first childbirth	0.63	0.61	0.61	0.60	0.58	0.59	0.58	0.6
Low age	0.60	0.61	0.62	0.61	0.60	0.59	0.57	0.58
Low BMI	0.59	0.59	0.59	0.59	0.57	0.57	0.57	0.58
Low length	0.59	0.60	0.62	0.62	0.61	0.61	0.59	0.60
Low menarche	0.60	0.62	0.62	0.62	0.60	0.61	0.58	0.59
Low weight	0.54	0.57	0.57	0.58	0.57	0.56	0.56	0.57
Ovarian cancer in family	0.37	0.52	0.48	0.57	0.61	0.61	0.57	0.59
Postmenopausal	0.66	0.64	0.64	0.64	0.63	0.63	0.61	0.62
Current smoker	0.59	0.63	0.61	0.60	0.60	0.59	0.59	0.60

[1] At least monthly use of alcohol. [2] Lifestyle/familial-expanded AI risk model. HRT—hormone replacement therapy.

Table A3. Age adjusted discriminatory performance of the lifestyle/familial-expanded AI risk model compared with the imaging-only AI risk model by cancer subtype. Table of 36 significant risk model prediction performance differences when comparing the lifestyle/familial-expanded AI risk model with the imaging-only AI risk model. The examined subtypes were mode of detection (screen-detected, symptomatic), stage (stage 1 or earlier, stage 2 or later), invasiveness (invasive, in situ), and receptor status (ER, PR, HER2). The significant differences are presented by years of follow-up and subtype. A *p*-value below 0.05 was considered a significant difference between the two risk model AUCs after adjusting the Holm–Bonferroni method to account for multiple comparisons.

Year	Cancer Subtype	Comparison AUC	Reference AUC	AUC Difference	Holm *p*-Value	Raw *p*-Value
2	ER	0.748	0.711	0.037	0.003	<0.001
2	Invasive	0.752	0.714	0.038	0.001	<0.001
2	PR	0.761	0.718	0.043	0.002	<0.001
2	Screen-detected	0.757	0.721	0.035	0.019	<0.001
3	ER	0.745	0.708	0.037	<0.001	<0.001
3	Invasive	0.745	0.707	0.038	<0.001	<0.001
3	PR	0.757	0.716	0.041	<0.001	<0.001
3	Screen-detected	0.753	0.719	0.034	<0.001	<0.001
4	ER	0.719	0.686	0.033	<0.001	<0.001
4	Symptomatic	0.706	0.679	0.027	0.001	<0.001
4	Invasive	0.713	0.677	0.035	<0.001	<0.001
4	PR	0.721	0.684	0.038	<0.001	<0.001
4	Screen-detected	0.736	0.700	0.036	<0.001	<0.001
5	ER	0.712	0.675	0.037	<0.001	<0.001
5	Symptomatic	0.705	0.673	0.033	<0.001	<0.001
5	Invasive	0.707	0.669	0.037	<0.001	<0.001
5	PR	0.713	0.675	0.038	<0.001	<0.001
5	Screen-detected	0.718	0.682	0.036	<0.001	<0.001
6	ER	0.705	0.667	0.038	<0.001	<0.001
6	Symptomatic	0.698	0.662	0.036	<0.001	<0.001
6	Invasive	0.700	0.662	0.039	<0.001	<0.001
6	PR	0.705	0.666	0.039	<0.001	<0.001
6	Screen-detected	0.714	0.679	0.035	<0.001	<0.001
8	ER	0.679	0.643	0.036	<0.001	<0.001
8	Symptomatic	0.679	0.644	0.035	<0.001	<0.001
8	Invasive	0.676	0.640	0.036	<0.001	<0.001
8	PR	0.680	0.644	0.036	<0.001	<0.001
8	Screen-detected	0.684	0.651	0.033	<0.001	<0.001
8	Stage 2 or later	0.707	0.677	0.030	<0.001	<0.001
10	ER	0.678	0.642	0.036	<0.001	<0.001
10	HER2	0.660	0.622	0.037	<0.001	<0.001
10	Symptomatic	0.678	0.649	0.030	<0.001	<0.001
10	Invasive	0.673	0.637	0.036	<0.001	<0.001
10	PR	0.678	0.643	0.034	<0.001	<0.001
10	Screen-detected	0.674	0.637	0.037	<0.001	<0.001
10	Stage 2 or later	0.706	0.683	0.023	<0.001	<0.001

ER—estrogen-receptor-positive, PR—progesterone-receptor-positive, HER2—human epidermal growth factor receptor 2-positive.

Table A4. Age adjusted discriminatory performance of Tyrer–Cuzick v8 compared with the imaging-only AI risk model by cancer subtype. Table of 36 significant risk model prediction performance differences when comparing the Tyrer–Cuzick v8 risk model with the imaging-only AI risk model. The examined subtypes were mode of detection (screen-detected, symptomatic), stage (stage 1 or earlier, stage 2 or later), invasiveness (invasive, in situ), and receptor status (ER, PR, HER2). The significant differences are presented by years of follow-up and subtype. A *p*-value below 0.05 was considered a significant difference between the two risk model AUCs after adjusting the Holm–Bonferroni method to account for multiple comparison.

Year	Cancer Subtype	Comparison AUC	Reference AUC	AUC Difference	Holm *p*-Value	Raw *p*-Value
1	Stage 2 or later	0.528	0.785	−0.258	0.021	<0.001
2	ER	0.613	0.711	−0.098	0.011	<0.001

Table A4. *Cont.*

Year	Cancer Subtype	Comparison AUC	Reference AUC	AUC Difference	Holm *p*-Value	Raw *p*-Value
2	In situ	0.631	0.801	−0.171	0.007	<0.001
2	Invasive	0.620	0.714	−0.095	0.015	<0.001
2	Screen-detected	0.606	0.721	−0.115	0.005	<0.001
2	Stage 2 or later	0.620	0.791	−0.171	0.01	<0.001
3	ER	0.618	0.708	−0.09	0.004	<0.001
3	In situ	0.622	0.772	−0.15	0.003	<0.001
3	Invasive	0.626	0.707	−0.081	0.003	<0.001
3	PR	0.628	0.716	−0.088	0.008	<0.001
4	ER	0.621	0.686	−0.065	0.046	<0.001
4	In situ	0.642	0.761	−0.119	0.035	<0.001
4	Invasive	0.617	0.677	−0.061	0.009	<0.001
4	Screen-detected	0.605	0.700	−0.095	<0.001	<0.001
4	Stage 2 or later	0.629	0.735	−0.106	0.007	<0.001
5	ER	0.612	0.675	−0.063	0.003	<0.001
5	In situ	0.603	0.726	−0.124	0.004	<0.001
5	Invasive	0.608	0.669	−0.061	0.001	<0.001
5	Screen-detected	0.590	0.682	−0.092	<0.001	<0.001
5	Stage 2 or later	0.631	0.726	−0.095	0.004	<0.001
6	ER	0.611	0.667	−0.056	0.018	<0.001
6	In situ	0.597	0.730	−0.133	<0.001	<0.001
6	Invasive	0.608	0.662	−0.054	0.001	<0.001
6	Screen-detected	0.586	0.679	−0.094	<0.001	<0.001
6	Stage 2 or later	0.631	0.701	−0.07	0.047	<0.001
8	ER	0.599	0.643	−0.043	0.024	<0.001
8	In situ	0.584	0.700	−0.116	<0.001	<0.001
8	Invasive	0.595	0.640	−0.045	0.005	<0.001
8	Screen-detected	0.575	0.651	−0.076	<0.001	<0.001
8	Stage 2 or later	0.609	0.677	−0.069	0.012	<0.001
10	In situ	0.591	0.673	−0.082	0.012	<0.001
10	invasive	0.600	0.637	−0.037	0.021	<0.001
10	Screen-detected	0.590	0.637	−0.047	0.018	<0.001
10	Stage 2 or later	0.613	0.683	−0.069	0.002	<0.001

ER—estrogen-receptor-positive, PR—progesterone-receptor-positive, HER2—human epidermal growth factor receptor 2-positive.

Table A5. Risk classification and stratification following the NICE guidelines of absolute risks at study entry for cases and the subcohort in the 10-year follow-up study population using the lifestyle/familial-expanded AI risk model, the imaging-only AI risk model, and the Tyrer–Cuzick v8 risk model.

Risk Group [1]	Cases, N (%)	Subcohort, N (%)	Absolute Risk, % [2]	Risk Ratio [3]
Expanded AI risk [4]				
Low	72/1661 (4.3%)	920/8110 (11%)	0.11	0.31
General	934/1661 (56%)	5748/8110 (71%)	0.36	1.0 (ref.)
Moderate	284/1661 (17%)	844/8110 (10%)	1.12	3.1
High	371/1661 (22%)	590/8110 (7.3%)	2.40	6.7
Imaging-only AI risk				
Low	74/1661 (4.5%)	743/8110 (9.2%)	0.12	0.32
General	1012/1661 (61%)	6047/8110 (75%)	0.37	1.0 (ref.)
Moderate	261/1661 (16%)	800/8110 (9.9%)	1.12	3.0
High	314/1661 (19%)	512/8110 (6.3%)	2.36	6.4
Tyrer–Cuzick v8				
Low	1/1661 (<0.1%)	20/8110 (0.2%)	0.68	0.29
General	1145/1661 (69%)	6500/8110 (80%)	2.30	1.0 (ref.)
Moderate	395/1661 (24%)	1381/8110 (17%)	5.05	2.2

Table A5. *Cont.*

Risk Group [1]	Cases, N (%)	Subcohort, N (%)	Absolute Risk, % [2]	Risk Ratio [3]
High	120/1661 (7.2%)	209/8110 (2.6%)	9.76	4.2

[1] The NICE guidelines 10-year absolute risk categories were general, moderate, and high using cut-offs on the absolute risk score. The cut-off between general and moderate risk was 3% for women below age 50 and 5% for women in age 50 and above. The cut-off between general and high risk was 8% for all women. A low-risk category was added using cut-off 0.75%. For the imaging-only AI risk model and the expanded AI risk model the risk cut-offs were adapted to 2-year risks by dividing the 10-year risk cut-offs by 5. [2] Median absolute risks of the subcohort are presented using the two-year risk scale for the two AI-based models and the ten-year risk scale for Tyrer–Cuzick v8. [3] The risk ratio was calculated as ratios of the absolute risks in each absolute risk category. [4] Lifestyle/familial-expanded AIrisk model. Tests were performed to determine the difference between the binomial proportions of breast cancers captured at high risk using the imaging-only AI risk model, the lifestyle/familial-expanded AI risk model, and Tyrer–Cuzick v8. The proportion of cancer captured by the lifestyle/familial-expanded AI risk model (22%) was significantly higher than the 19% captured by the image-based risk model, $p = 0.01$, which in turn was significantly higher than the 7.2% of breast cancers captured by Tyrer–Cuzick v8, $p < 0.01$. The risk ratio between high-risk and low-risk was 21.7 for the lifestyle/familial expanded AI risk model, 20.3 for the imaging-only AI risk model, and 14.6 for Tyrer–Cuzick v8. NICE—National Institute for Health and Care Excellence.

Table A6. Risk classification per USPSTF guidelines of absolute risks at enrollment for breast cancers and women in the subcohort using three risk models. Risk classification and stratification following the USPSTF guidelines of absolute risks at study entry for cases and the subcohort in the 10-year follow-up study population using the lifestyle/familial-expanded AI risk model, the imaging-only AI risk model, and Tyrer–Cuzick v8.

Risk Group [1]	Cases, N (%)	Subcohort, N (%)	Absolute Risk, % [2]	Risk Ratio [3]
Expanded AI risk [4]				
General	189/1661 (11%)	2180/8110 (27%)	0.16	1.0 (ref.)
Moderate	960/1661 (58%)	5025/8110 (62%)	0.46	2.8
High	512/1661 (31%)	905/8110 (11%)	1.90	11.7
Imaging-only risk				
General	208/1661 (13%)	1948/8110 (24%)	0.17	1.0 (ref.)
Moderate	1025/1661 (62%)	5334/8110 (66%)	0.45	2.6
High	428/1661 (26%)	828/8110 (10%)	1.83	10.7
Tyre–Cuzick v8				
General	25/1661 (1.5%)	300/8110 (3.7%)	1.05	1.0 (ref.)
Moderate	1365/1661 (82%)	7248/8110 (89%)	2.53	2.4
High	271/1661 (16%)	562/8110 (6.9%)	7.51	7.2

[1] The USPSTF guidelines' 5-year absolute risk categories were general, moderate, and high. The cut-off for high-risk is based on 3% 5-year risk and the cut-off for low-risk is based on 0.6% 5-year risk, that is, the average risk of a 40-year-old woman. For the imaging-only AI risk model and the expanded AI risk model, the cut-offs were adapted to 2-year risks by dividing the 5-year risk cut-offs by 2.5. [2] Median absolute risks are presented using the two-year risk scale for the two AI-based models and the ten-year risk scale for Tyrer–Cuzick v8. [3] The risk ratio was calculated as ratios of the absolute risks in each absolute risk category. [4] Lifestyle/familial-expanded AI risk model. Tests were performed to determine the difference between the binomial proportions of breast cancers captured at high risk by the lifestyle/familial-expanded AI risk model, the imaging-only AI risk model, and Tyrer–Cuzick v8. The proportion of cancer captured by the lifestyle/familial-expanded AI risk model (31%) was significantly higher than the 26% captured by the image-based risk model, $p = 0.01$, which in turn was significantly higher than the 16% of breast cancers captured by Tyrer–Cuzick v8 10-year, $p < 0.01$. USPSTF—U.S. Preventive Service Task Force.

References

1. Gail, M.H.; Pfeiffer, R.M. Breast Cancer Risk Model Requirements for Counseling, Prevention, and Screening. *J. Natl. Cancer Inst.* **2018**, *110*, 994–1002. [CrossRef] [PubMed]
2. Britt, K.L.; Cuzick, J.; Phillips, K.A. Key steps for effective breast cancer prevention. *Nat. Rev. Cancer* **2020**, *20*, 417–436. [CrossRef] [PubMed]
3. Yala, A.; Mikhael, P.G.; Strand, F.; Lin, G.; Smith, K.; Wan, Y.-L.; Lamb, L.; Hughes, K.; Lehman, C.; Barzilay, R. Toward robust mammography-based models for breast cancer risk. *Sci. Transl. Med.* **2021**, *13*, eaba4373. [CrossRef]
4. Eriksson, M.; Destounis, S.; Czene, K.; Zeiberg, A.; Day, R.; Conant, E.F.; Schilling, K.; Hall, P. A risk model for digital breast tomosynthesis to predict breast cancer and guide clinical care. *Sci. Transl. Med.* **2022**, *14*, eabn3971. [CrossRef]

5.   Louro, J.; Posso, M.; Hilton Boon, M.; Roman, M.; Domingo, L.; Castells, X.; Sala, M. A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br. J. Cancer* **2019**, *121*, 76–85. [CrossRef]

6.   Smith, R.A.; Andrews, K.S.; Brooks, D.; Fedewa, S.A.; Manassaram-Baptiste, D.; Saslow, D.; Wender, R.C. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J. Clin.* **2019**, *69*, 184–210. [CrossRef] [PubMed]

7.   Bevers, T.B.; Ward, J.H.; Arun, B.K.; Colditz, G.A.; Cowan, K.H.; Daly, M.B.; Garber, J.E.; Gemignani, M.L.; Gradishar, W.J.; Jordan, J.A.; et al. Breast Cancer Risk Reduction, Version 2.2015. *J. Natl. Compr. Cancer Netw.* **2015**, *13*, 880–915. [CrossRef]

8.   Monticciolo, D.L.; Newell, M.S.; Hendrick, R.E.; Helvie, M.A.; Moy, L.; Monsees, B.; Kopans, D.B.; Eby, P.R.; Sickles, E.A. Breast Cancer Screening for Average-Risk Women: Recommendations From the ACR Commission on Breast Imaging. *J. Am. Coll. Radiol.* **2017**, *14*, 1137–1143. [CrossRef]

9.   Cuzick, J.; Sestak, I.; Cawthorn, S.; Hamed, H.; Holli, K.; Howell, A.; Forbes, J.F.; Investigators, I.-I. Tamoxifen for prevention of breast cancer: Extended long-term follow-up of the IBIS-I breast cancer prevention trial. *Lancet Oncol.* **2015**, *16*, 67–75. [CrossRef]

10.   Fisher, B.; Costantino, J.P.; Wickerham, D.L.; Cecchini, R.S.; Cronin, W.M.; Robidoux, A.; Bevers, T.B.; Kavanah, M.T.; Atkins, J.N.; Margolese, R.G.; et al. Tamoxifen for the prevention of breast cancer: Current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J. Natl. Cancer Inst.* **2005**, *97*, 1652–1662. [CrossRef]

11.   Powles, T.J.; Ashley, S.; Tidy, A.; Smith, I.E.; Dowsett, M. Twenty-year follow-up of the Royal Marsden randomized, double-blinded tamoxifen breast cancer prevention trial. *J. Natl. Cancer Inst.* **2007**, *99*, 283–290. [CrossRef] [PubMed]

12.   Chen, Z.; Xu, L.; Shi, W.; Zeng, F.; Zhuo, R.; Hao, X.; Fan, P. Trends of female and male breast cancer incidence at the global, regional, and national levels, 1990–2017. *Breast Cancer Res. Treat.* **2020**, *180*, 481–490. [CrossRef] [PubMed]

13.   Mocellin, S.; Goodwin, A.; Pasquali, S. Risk-reducing medications for primary breast cancer: A network meta-analysis. *Cochrane Database Syst. Rev.* **2019**, *4*, CD012191. [CrossRef]

14.   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 CG164*; NCCC: London, UK, 2013; ISBN 978-1-4731-0160-9. Available online: https://www.nice.org.uk/guidance/cg164 (accessed on 3 February 2023).

15.   Siu, A.L.; US Preventive Services Task Force. Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Ann. Intern. Med.* **2016**, *164*, 279–296. [CrossRef] [PubMed]

16.   Gabrielson, M.; Eriksson, M.; Hammarstrom, M.; Borgquist, S.; Leifland, K.; Czene, K.; Hall, P. Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). *Int. J. Epidemiol.* **2017**, *46*, 1740–1741g. [CrossRef]

17.   Eriksson, M.; Czene, K.; Strand, F.; Zackrisson, S.; Lindholm, P.; Lång, K.; Förnvik, D.; Sartor, H.; Mavaddat, N.; Easton, D. Identification of women at high risk of breast cancer who need supplemental screening. *Radiology* **2020**, *297*, 327–333. [CrossRef]

18.   iCADMedical. ICAD Profound AI Risk. Available online: https://icadmed.com/profoundai-risk.html (accessed on 3 May 2023).

19.   Zidar, M.N.; Larm, P.; Tillgren, P.; Akhavan, S. Non-attendance of mammographic screening: The roles of age and municipality in a population-based Swedish sample. *Int. J. Equity Health* **2015**, *14*, 157. [CrossRef]

20.   Eriksson, M.; Li, J.; Leifland, K.; Czene, K.; Hall, P. A comprehensive tool for measuring mammographic density changes over time. *Breast Cancer Res. Treat.* **2018**, *169*, 371–379. [CrossRef]

21.   Tyrer, J.; Duffy, S.W.; Cuzick, J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat. Med.* **2004**, *23*, 1111–1130. [CrossRef]

22.   Yang, X.; Eriksson, M.; Czene, K.; Lee, A.; Leslie, G.; Lush, M.; Wang, J.; Dennis, J.; Dorling, L.; Carvalho, S.; et al. Prospective validation of the BOADICEA multifactorial breast cancer risk prediction model in a large prospective cohort study. *J. Med. Genet.* **2022**, *59*, 1196–1205. [CrossRef]

23.   Eriksson, M.; Czene, K.; Pawitan, Y.; Leifland, K.; Darabi, H.; Hall, P. A clinical model for identifying the short-term risk of breast cancer. *Breast Cancer Res.* **2017**, *19*, 29. [CrossRef] [PubMed]

24.   Ehinger, A.; Lindman, H.; Löfgren, L.; Petersson, L.-M.; Sandelin, K. *Årsrapport 2015 Från Nationella Bröstcancerregistret*; Regionalt Cancercentrum Stockholm Gotland: Stockholm, Sweden, 2016.

25.   Amin, M.B.; Greene, F.L.; Edge, S.B.; Compton, C.C.; Gershenwald, J.E.; Brookland, R.K.; Meyer, L.; Gress, D.M.; Byrd, D.R.; Winchester, D.P. The eighth edition AJCC cancer staging manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA A Cancer J. Clin.* **2017**, *67*, 93–99. [CrossRef] [PubMed]

26.   Sharp, S.J.; Poulaliou, M.; Thompson, S.G.; White, I.R.; Wood, A.M. A review of published analyses of case-cohort studies and recommendations for future reporting. *PLoS ONE* **2014**, *9*, e101176. [CrossRef] [PubMed]

27.   Barlow, W.E.; Ichikawa, L.; Rosner, D.; Izumi, S. Analysis of case-cohort designs. *J. Clin. Epidemiol.* **1999**, *52*, 1165–1172. [CrossRef] [PubMed]

28.   Janes, H.; Pepe, M.S. Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika* **2009**, *96*, 371–382. [CrossRef]

29.   De Carvalho, V.I.; Rodriguez-Alvarez, M.X. Bayesian nonparametric inference for the covariate-adjusted ROC curve. *arXiv* **2018**, arXiv:1806.00473.

30.   Carpenter, J.; Bithell, J. Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Stat. Med.* **2000**, *19*, 1141–1164. [CrossRef]

31.   Gibson, G. Hints of hidden heritability in GWAS. *Nat. Genet.* **2010**, *42*, 558–560. [CrossRef]

32. Viegas, F.B.; Wattenberg, M.; Van Ham, F.; Kriss, J.; McKeon, M. Manyeyes: A site for visualization at internet scale. *IEEE Trans. Vis. Comput. Graph.* **2007**, *13*, 1121–1128. [CrossRef]

33. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.

34. Altman, D.G.; Bland, J.M. Statistics Notes: Diagnostic tests 2: Predictive values. *BMJ* **1994**, *309*, 102. [CrossRef]

35. The R Development Core Team. *R: A Language and Environment for Statistical Computing*; The R Development Core Team: Vienna, Austria, 2013.

36. Dumas-Mallet, E.; Button, K.S.; Boraud, T.; Gonon, F.; Munafo, M.R. Low statistical power in biomedical science: A review of three human research domains. *R. Soc. Open Sci.* **2017**, *4*, 160254. [CrossRef]

37. Niell, B.L.; Augusto, B.; McIntyre, M.; Conley, C.C.; Gerke, T.; Roetzheim, R.; Garcia, J.; Vadaparampil, S.T. Distribution of Estimated Lifetime Breast Cancer Risk Among Women Undergoing Screening Mammography. *Am. J. Roentgenol.* **2021**, *217*, 48–55. [CrossRef] [PubMed]

38. Saghatchian, M.; Abehsera, M.; Yamgnane, A.; Geyl, C.; Gauthier, E.; Hélin, V.; Bazire, M.; Villoing-Gaudé, L.; Reyes, C.; Gentien, D. Feasibility of personalized screening and prevention recommendations in the general population through breast cancer risk assessment: Results from a dedicated risk clinic. *Breast Cancer Res. Treat.* **2022**, *192*, 375–383. [CrossRef] [PubMed]

39. McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.S.; Darzi, A.; et al. International evaluation of an AI system for breast cancer screening. *Nature* **2020**, *577*, 89–94. [CrossRef] [PubMed]

40. Socialstyrelsen. *Screening för Bröstcancer: Rekommendation Och Bedömningsunderlag*; Artikelnr: 2014-2-32; Socialstyrelsen: Stockholm, Sweden, 2014.