



OPEN ACCESS

Original research

Prospective validation of the BOADICEA multifactorial breast cancer risk prediction model in a large prospective cohort study

Xin Yang ,¹ Mikael Eriksson,² Kamila Czene,² Andrew Lee ,¹ Goska Leslie,¹ Michael Lush,¹ Jean Wang,¹ Joe Dennis,¹ Leila Dorling,¹ Sara Carvalho,¹ Nasim Mavaddat,¹ Jacques Simard,³ Marjanka K Schmidt,^{4,5} Douglas F Easton,^{1,6} Per Hall,^{2,7} Antonis C Antoniou ¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jmg-2022-108806>).

For numbered affiliations see end of article.

Correspondence to

Professor Antonis C Antoniou, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge CB1 8RN, Cambridgeshire, UK; aca20@medschl.cam.ac.uk

Received 5 July 2022

Accepted 24 August 2022

Published Online First 26 September 2022



► <http://dx.doi.org/10.1136/jmedgenet-2022-108471>



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY. Published by BMJ.

To cite: Yang X, Eriksson M, Czene K, et al. *J Med Genet* 2022;**59**:1196–1205.

ABSTRACT

Background The multifactorial Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) breast cancer risk prediction model has been recently extended to consider all established breast cancer risk factors. We assessed the clinical validity of the model in a large independent prospective cohort.

Methods We validated BOADICEA (V.6) in the Swedish KARolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA) cohort including 66 415 women of European ancestry (median age 54 years, IQR 45–63; 816 incident breast cancers) without previous cancer diagnosis. We calculated 5-year risks on the basis of questionnaire-based risk factors, pedigree-structured first-degree family history, mammographic density (BI-RADS), a validated breast cancer polygenic risk score (PRS) based on 313-SNPs, and pathogenic variant status in 8 breast cancer susceptibility genes: *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1*. Calibration was assessed by comparing observed and expected risks in deciles of predicted risk and the calibration slope. The discriminatory ability was assessed using the area under the curve (AUC).

Results Among the individual model components, the PRS contributed most to breast cancer risk stratification. BOADICEA was well calibrated in predicting the risks for low-risk and high-risk women when all, or subsets of risk factors are included in the risk prediction. Discrimination was maximised when all risk factors are considered (AUC=0.70, 95% CI: 0.66 to 0.73; expected-to-observed ratio=0.88, 95% CI: 0.75 to 1.04; calibration slope=0.97, 95% CI: 0.95 to 0.99). The full multifactorial model classified 3.6% women as high risk (5-year risk $\geq 3\%$) and 11.1% as very low risk (5-year risk $< 0.33\%$).

Conclusion The multifactorial BOADICEA model provides valid breast cancer risk predictions and a basis for personalised decision-making on disease prevention and screening.

INTRODUCTION

Breast cancer (BC) is the most common cancer diagnosed among women worldwide and is associated with significant mortality.¹ Screening and prevention options are available, including mammography, MRI, chemoprevention and risk-reducing surgery,

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ No study has assessed the clinical validity of the full multifactorial Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) model for predicting future breast cancer risks.

WHAT THIS STUDY ADDS

⇒ This is the first study to validate the comprehensive BOADICEA model based on the joint effects of family history, questionnaire-based risk factors, mammographic density, polygenic risk score and rare pathogenic variants in all eight established breast cancer susceptibility genes.
⇒ The model is well calibrated overall and predicts risks accurately in different categories of predicted risk.
⇒ It discriminates well between affected and unaffected women and can result in clinically meaningful levels of breast cancer risk stratification.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ BOADICEA is freely available via the CanRisk tool (www.canrisk.org) and has been incorporated in several clinical management guidelines in the UK, North America and other countries.
⇒ BOADICEA can be used by healthcare professionals in personalising risk assessment to facilitate shared decision-making with patients on lifestyle changes, prevention or screening options for managing breast cancer risk.
⇒ It can be used to identify high-risk women who may benefit most from enhanced screening or other preventive or risk-reducing treatments, and also to identify low-risk women who are unlikely to benefit from such interventions, which are also associated with adverse effects.
⇒ The results are based mainly on data from women of European ancestry. Further model customisation and validation is needed before applying the model to women from other ancestry groups.

but these are costly and may be associated with overdiagnosis, overtreatment or adverse effects.²⁻⁴ BC risk prediction models have the potential of improving individualised BC risk assessment and population risk stratification.⁵ They can be used for identifying women at high risk who are most likely to benefit from such interventions.

The Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) model as implemented in the CanRisk tool (www.canrisk.org) is currently used by clinicians to estimate the future risk of developing BC.⁶⁻¹¹ The latest V.6 models BC risk using seven types of risk factors¹²⁻¹³: (1) pathogenic variants (PVs) in eight high-risk or moderate-risk BC susceptibility genes *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1*; (2) a validated polygenic risk score (PRS), using 313 common BC genetic susceptibility variants¹⁴; (3) detailed cancer family history (FH), including age at cancer diagnosis for affected relatives and age at last follow-up or death for unaffected relatives; (4) a residual polygenic component accounting for familial aggregation not explained by the above observed genetic effects; (5) nine questionnaire-based lifestyle, hormonal and reproductive risk factors (QRFs); (6) mammographic density (MD) measured using the BI-RADS breast composition categorisation¹⁵ and (7) demographic factors including age, year of birth and country. This is the first comprehensive BC risk prediction model considering all established BC risk factors, MD, PRS and PVs in all eight established BC susceptibility genes and is hypothesised to improve risk stratification compared with older versions of BOADICEA that considered only FH and *BRCA1/2* PV information.¹⁶⁻¹⁷

Previous prospective validation studies have been either small or were based on women ascertained in high-risk settings; assessed older versions of BOADICEA, based only on FH and *BRCA1/2* PV information or used only a subset of the risk factors considered in the model.¹⁸⁻²¹ None assessed the full multifactorial model. Here, for the first time, we evaluate the performance of the full BOADICEA model¹²⁻¹³ in an independent population-based prospective cohort of >60 000 women.

METHODS

Study participants

The KARolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA) cohort is a large prospective population-based cohort of women who were invited to the Swedish national mammography screening programme between January 2011 and March 2013.²² In the Swedish national screening programme, women are invited from age 40 to 74 years. Approximately 3% of the women are recalled following a screening mammogram for further investigation due to a suspicious finding. Additionally, approximately 0.2% of the women have a mammogram between scheduled screens due to symptoms or clinical referral. The present study included regularly screened women and symptomatic women who had mammograms performed on screening machines at study baseline at four hospitals in Sweden. The women were followed for BC events until 2017 with a mean/median follow-up of 5.7/6.0 years.²² All participants completed a baseline questionnaire, which included anthropometric/lifestyle/hormonal and reproductive factors. Full-field digital processed mammograms of right and left breasts were collected at baseline to measure MD using the area-based STRATUS software, which estimated the MD on digital processed mammograms using a machine learning method.²³ The percentage of MD was then calculated as the ratio of the dense area to the total area, which were categorised into the

four computer-assessed BI-RADS categories by fixed thresholds (<2%, 2-8%, 8-49% and >49%).¹⁵⁻²³⁻²⁴ The resulting age-specific BI-RADS distributions were in line with data from the general population²⁵ and used in BOADICEA, although there is a somewhat higher proportion of women in category 'A' in particular among women aged 50 years or older (online supplemental table S1). Self-reported FH data were coded as pedigrees, including information on ages at BC diagnosis or last observation for first-degree relatives. Personal cancer diagnosis and death were obtained through linkages with high-quality healthcare registers, which captured 98.5% of all incident cancers within 12 months of diagnosis.²²⁻²⁶ For this analysis, the last linkage was performed in June 2017. Each participant's follow-up was censored at date of BC (invasive or in situ), bilateral prophylactic mastectomy, last linkage date, death, baseline plus 5 years or age 80 years, whichever occurred first. We restricted our analyses to women with no previous cancer (any) or risk-reducing mastectomy history who had information on FH, QRFs and MD.

Genetic data

Genotyping and gene-panel sequencing data were available for a subset of the cohort based on previous genotyping experiments.²⁷⁻²⁸ Genotyping for the 313-SNP PRS¹⁴ was performed on the majority of the BC incident cases available and randomly selected unaffected women from the entire cohort that reflected the distribution in the entire KARMA cohort. A subset of the participants with PRS was sequenced for a 34-gene panel through the Breast Cancer Risk after Diagnostic GENE Sequencing (BRIDGES) study²⁷ using the same sampling approach. For this subcohort, PV carrier status was determined in the *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1* genes. The risk factor distributions for all QRFs (including any missing categories) were virtually the same in the entire KARMA cohort and the two subcohorts with genetic data (online supplemental table S2).

Risk prediction

We predicted 5-year BC risks using BOADICEA V.6,¹² with Swedish age-specific and calendar period-specific population incidences for invasive BC. To exclude potentially prevalent patients with undiagnosed BC at study recruitment, we predicted the 5-year BC risks starting from the age at baseline plus 1 year. Women were considered as affected only if they developed invasive BC within this 5-year risk prediction interval. For unaffected women with follow-up time shorter than 5 years, the BC risks were predicted to the censored age. Since BOADICEA allows for missing questionnaire-based risk factor information,¹² risk predictions were carried out for all women irrespective of individual missing QRFs. Predictions involving genetic data were carried out only in the subcohorts of women with the genetic data available (see "Statistical analysis").

Statistical analysis

Model calibration was assessed by the ratio of the expected (E) to the observed (O) BC risk in the dataset.²⁹⁻³⁰ We also assessed the agreement between predicted and observed risk for each individual using the calibration slope, which was calculated by fitting a logistic regression where the dependent variable was the observed outcome (affected/unaffected) and the independent variable was the log-odds of the predicted risk. The calibration slope assesses whether the predicted risks are too extreme or conversely too moderate, especially at the high-risk and low-risk tails and is expected to be equal to 1 if the model is perfectly

calibrated. The observed and expected risks were also compared in decile categories of predicted risks or by menopausal status at recruitment. Discrimination was assessed by the area under the receiver operating characteristic curve (AUC) and Harrell's concordance (C) index.^{31–33} Bootstrap (100 replications) was used to obtain CIs for Harrell's C-index estimates.^{31 32}

We assessed the contribution of individual and combinations of risk factors in predicting BC risk. All participants had information on age, year of birth, QRFs, FH, MD and were from Sweden. Of those, 15 502 participants had also information on PRS and 5693 had PRS and gene-panel sequencing data. When assessing the model performance in the subcohorts, to avoid potential biases we considered the explicit sampling scheme by using an inverse probability weighted approach,^{21 29 34} which reflects the probability of participants being included in the subcohort. The sample inclusion probabilities were computed by fitting a logistic regression model in the full dataset in which the outcome (inclusion or not) was dependent on the age at baseline, follow-up duration, incident BC status and the interaction between BC status, age at baseline and the follow-up duration.³⁴ The weights were then the inverse of the fitted inclusion probabilities for each individual. To allow for a direct comparison of the models with or without genetic factors, model comparison were performed in the subcohorts with genetic information.

To evaluate the potential improvements in risk classification when adding QRFs, MD, PRS and gene-panel sequencing data, to FH (previous version of BOADICEA), we calculated the proportion of all participants and incident BCs reclassified as high risk. Women with 5-year BC risks above 3% were considered as high risk, in line with the risk threshold recommended for risk-reducing treatments by the US Preventive Services Task Force.³⁵

All analyses were performed in R (V3.6.3)³⁶ and the R package 'iCare' was used to calculate the E/O and AUC.²⁹ Sensitivities and specificities were calculated using R package 'epiR'.³⁷ All statistical tests were two-sided.

RESULTS

A total of 66 415 participants (aged 25–74 years) were eligible for inclusion in the analysis with information available on FH, QRFs and MD (figure 1, online supplemental table S2). Among these, 816 developed BC within the 5-year risk prediction horizon (tumour characteristics in online supplemental table S3); 90.1% were of self-reported European ancestry (online supplemental table S2). Of the eligible participants, 15 502 women (23.3%) had information on the 313-SNP PRS (676 affected) and 5693 women (8.6%) had PRS and PV status in *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1*, of whom 280 were affected (figure 1, online supplemental table S2). A detailed summary of the genetic and epidemiological characteristics of participants at baseline are shown in online supplemental table S2).

Model discrimination and calibration

Using the entire cohort, FH, QRFs or MD individually resulted in similar estimated AUCs and Harrell's C-indexes of 0.62 (table 1, online supplemental figure S1). The AUC was maximised when FH, QRFs and MD were considered jointly (AUC=0.64, 95%CI: 0.62 to 0.66). For this model, the overall ratio of expected to observed number of cases (E/O) was 0.92 (95% CI: 0.86 to 0.99) with a slight underestimation of risk in the middle deciles of predicted risk when MD was included (online supplemental figure S2). However, the model was well-calibrated in the

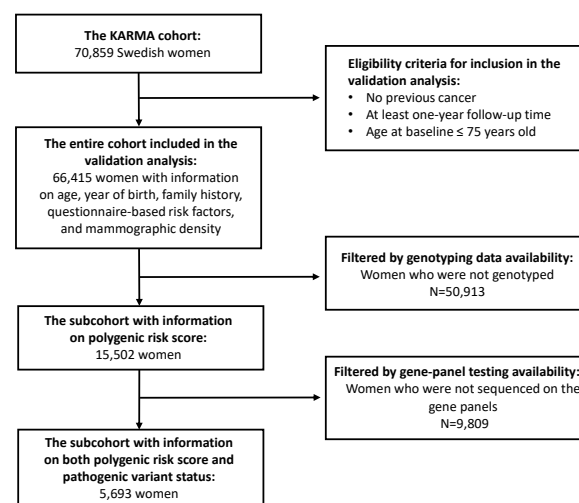


Figure 1 Consolidated Standards of Reporting Trials diagram summarising the KARMA cohort data.

bottom and top deciles of predicted risk. The calibration slope was 0.98 (95% CI: 0.96 to 1.00, table 1).

The estimated calibration and discrimination statistics for models that considered FH, QRFs or MD individually or jointly were similar in the weighted subcohort of women with PRS information and the full cohort (table 1), indicating no evidence of bias due to sampling. The PRS provided the widest distribution of predicted risks among the individual model components (online supplemental figure S3), and discriminated best between patients with incident BC and unaffected women (AUC=0.67, 95%CI: 0.64 to 0.69, table 1). The calibration slope for this model was 1.02 (95% CI: 1.00 to 1.03) and the overall E/O 1.06 (95% CI: 0.99 to 1.15). When the PRS was combined with FH and QRFs, both discrimination and calibration were improved (AUC=0.68, 95%CI: 0.66 to 0.70; calibration slope=1.00, 95% CI: 0.98 to 1.02; E/O=1.01, 95% CI: 0.94 to 1.09; figure 2, table 1). The addition of MD into the model further improved the discrimination (AUC=0.69, 95% CI: 0.67 to 0.71, table 1) but resulted in some risk underprediction (E/O=0.92, 95% CI: 0.85 to 0.99) mainly in the eighth decile of predicted risks, although there was no evidence for systematic underprediction (calibration slope=0.98, 95% CI: 0.96 to 1.00, figure 2).

Using the weighted subcohort of women with both PRS and PV information, the full model including FH, QRFs, MD, PRS and PVs predicted the widest BC risk distribution (online supplemental figure S4) and maximised model discrimination (AUC=0.70, 95%CI: 0.66 to 0.73, table 1). The calibration slope was 0.97 (95% CI: 0.95 to 0.99) and the overall E/O 0.88 (95% CI: 0.75 to 1.04, table 1, figure 3). The possible underprediction was primarily driven by the inclusion of MD (for the model without MD: E/O=1.00, 95% CI: 0.84 to 1.17, calibration slope=1.00, 95% CI: 0.98 to 1.02). However, the model was well calibrated in the bottom and high deciles of predicted risk (figure 3).

Table 1 Calibration and discrimination of 5-year predicted breast cancer risks under the BOADICEA model using different risk factor combinations

Model	AUC (95% CI)	Harrell's C-index (95% CI)	E/O (95% CI)	Calibration slope (95% CI)
Entire cohort with information on FH, QRFs and MD (n=66 415; n.BCs=816)				
FH	0.62 (0.60 to 0.64)	0.62 (0.60 to 0.64)	1.12 (1.05 to 1.20)	1.03 (1.01 to 1.05)
QRFs	0.62 (0.60 to 0.64)	0.62 (0.60 to 0.64)	0.95 (0.89 to 1.02)	0.99 (0.97 to 1.01)
MD	0.62 (0.60 to 0.64)	0.62 (0.60 to 0.64)	0.94 (0.88 to 1.01)	0.99 (0.97 to 1.01)
FH+QRFs	0.63 (0.62 to 0.65)	0.63 (0.61 to 0.64)	1.02 (0.95 to 1.09)	1.01 (0.99 to 1.02)
FH+QRFs+MD	0.64 (0.62 to 0.66)	0.64 (0.61 to 0.66)	0.92 (0.86 to 0.99)	0.98 (0.96 to 1.00)
Subcohort with information on FH, QRFs, MD and PRS (n=15 502; n.BCs=676)				
FH	0.61 (0.59 to 0.64)	0.63 (0.60 to 0.65)	1.12 (1.04 to 1.21)	1.03 (1.01 to 1.05)
QRFs	0.63 (0.61 to 0.65)	0.63 (0.61 to 0.65)	0.94 (0.87 to 1.01)	0.99 (0.97 to 1.01)
MD	0.63 (0.61 to 0.65)	0.64 (0.62 to 0.66)	0.93 (0.87 to 1.01)	0.99 (0.97 to 1.00)
PRS	0.67 (0.64 to 0.69)	0.67 (0.65 to 0.69)	1.06 (0.99 to 1.15)	1.02 (1.00 to 1.03)
FH+QRFs+MD	0.65 (0.63 to 0.67)	0.65 (0.63 to 0.67)	0.91 (0.84 to 0.98)	0.98 (0.96 to 1.00)
FH+QRFs+PRS	0.68 (0.66 to 0.70)	0.68 (0.66 to 0.70)	1.01 (0.94 to 1.09)	1.00 (0.98 to 1.02)
FH+QRFs+PRS+MD	0.69 (0.67 to 0.71)	0.69 (0.67 to 0.71)	0.92 (0.85 to 0.99)	0.98 (0.96 to 1.00)
Subcohort with information on FH, QRFs, MD, PRS and PV status (n=5693; n.BCs=280)				
FH+QRFs+PRS+MD	0.69 (0.64 to 0.73)	0.70 (0.67 to 0.73)	0.88 (0.74 to 1.04)	0.97 (0.95 to 0.99)
FH+QRFs+PRS+MD+PV	0.70 (0.66 to 0.73)	0.71 (0.68 to 0.74)	0.88 (0.75 to 1.04)	0.97 (0.95 to 0.99)

BC, breast cancer; BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; C-index, concordance index; E, expected number of BCs in the 5-year period; FH, family history; MD, mammographic density in BI-RADS; n.BCs, number of patients with incident breast cancer; O, observed number BCs; PRS, polygenic risk score; PV, pathogenic variants in *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1*; QRFs, questionnaire-based risk factors.

Model performance by menopausal status

Among women with PRS data, there were 5714 premenopausal (212 BCs) and 9788 postmenopausal women (464 BCs). When considering FH, QRFs and PRS, the model was well calibrated in both premenopausal (calibration slope=1.02, 95% CI: 0.99 to 1.05) and postmenopausal women (calibration slope=0.99, 95% CI: 0.97 to 1.02). The AUC for premenopausal women was 0.68 (95% CI: 0.64 to 0.71) compared with 0.66 (95% CI: 0.63

to 0.68) in postmenopausal women; table 2, figure 4A). When MD was added, the AUC was increased for both premenopausal (AUC=0.69, 95% CI: 0.65 to 0.72) and postmenopausal women (AUC=0.67, 95% CI: 0.64 to 0.69, table 2). However, there was some underprediction of risk among postmenopausal women, for the middle seventh-eighth deciles of predicted risk (figure 4). The overall calibration slopes in premenopausal and postmenopausal women were 1.03 (95% CI: 1.00 to 1.06) and

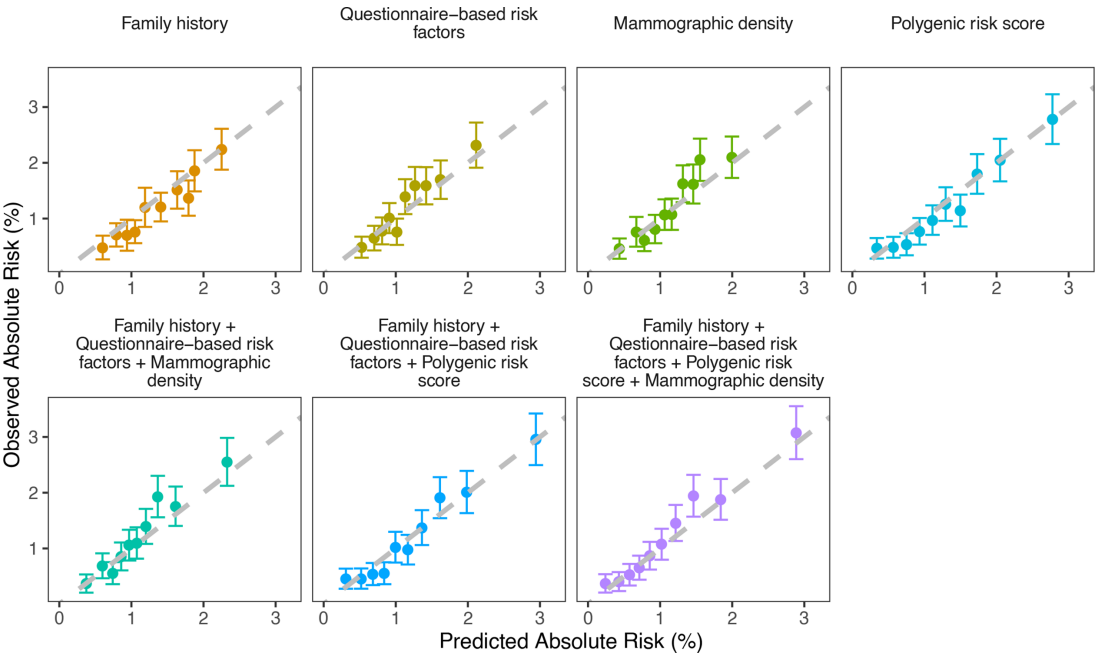


Figure 2 Observed and predicted 5-year breast cancer risks using the subcohort of participants with polygenic risk score information (n=15 502) under different risk factor combinations. Women were grouped into deciles of predicted risks. Each dot represents the mean observed and predicted risk in the decile and the vertical segments represent 95% CIs. The dashed line is the diagonal line with slope equal to 1 (corresponding to expected to observed number of cases ratio of 1 for each decile). When the CI crosses the diagonal, the decile-predicted risk is not significantly different from the observed risk. When a dot and the associated CI fall above the diagonal, there is a suggestion for underprediction of risk; when a dot and associated CI fall below the diagonal, there is a suggestion for overprediction of risk.

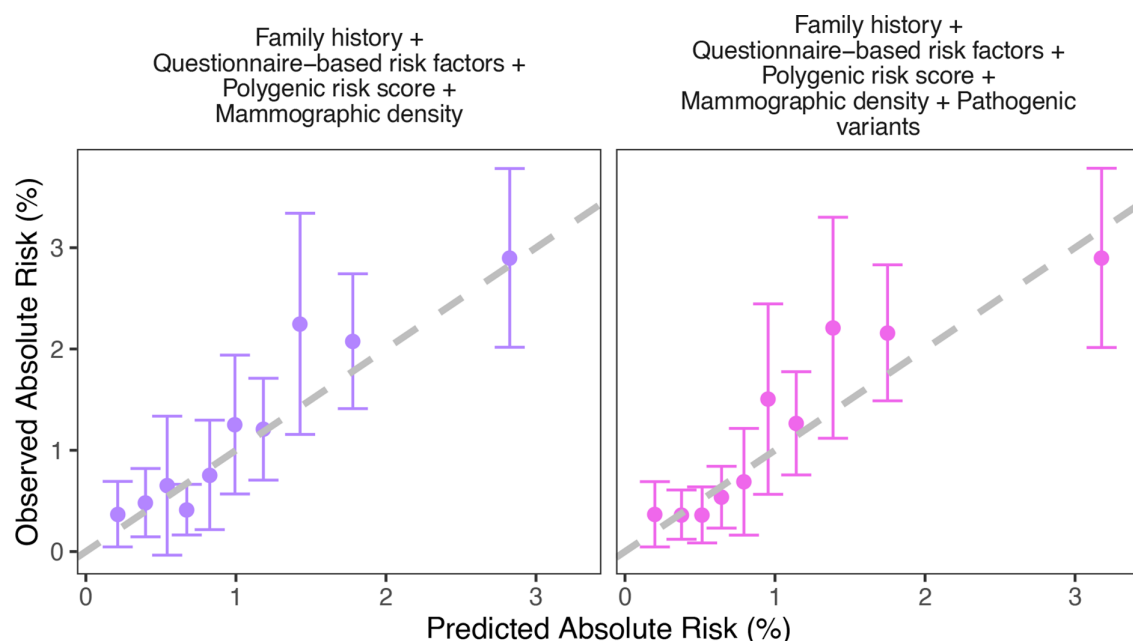


Figure 3 Observed and predicted 5-year breast cancer risks using the subcohort of participants with polygenic risk score and pathogenic variant status information (n=5 693) under different models. Women were grouped into deciles of predicted risks.

0.95 (95% CI: 0.93 to 0.97), respectively. Among women with PRS and PV information, there were 1293 premenopausal (77 incident cancers) and 4400 postmenopausal women (203 incident cancers). When considering the full multifactorial model including PVs, the AUC in premenopausal women was 0.69 (95% CI: 0.62 to 0.76) and in postmenopausal women 0.68 (95% CI: 0.64 to 0.73). There was an overall underprediction of risk among postmenopausal women under the full model (E/O=0.78, 95% CI: 0.65 to 0.93; calibration slope=0.94, 95% CI: 0.92 to 0.96, [table 2](#), [figure 4B](#)), but there was no significant difference between observed and predicted risk for any of the risk deciles. The full model without MD was well calibrated overall (E/O=0.94, 95% CI: 0.79 to 1.13; calibration slope=0.99, 95% CI: 0.96 to 1.01).

Risk classification, sensitivity/specificity

The model considering FH alone classified 0.8% women as high risk (5-year risk $\geq 3\%$), including 1.7% of the patients with incident BC. The addition of QRFs, PRS and MD reclassified 2.8%

women from the low-risk (risk $< 3\%$) to the high-risk (risk $\geq 3\%$) group and 0.5% in the opposite direction (online supplemental table S4). Under this model, the sensitivity was 69%, 38% and 9% at the $\geq 1.1\%$, $\geq 1.67\%$ and $\geq 3\%$ risk thresholds (online supplemental table S5); and was 37%, 53% and 65% when considering the top 20%, 30% and 40% of the predicted risks distribution, respectively (online supplemental table S6).

The addition of PV information reclassified 1.0% women from the low-risk to the high-risk group and 0.3% from the high-risk to the low-risk group (online supplemental table S7). Under this model, 3.6% women were classified as high risk, including 8.9% of the patients with incident BC during the 5-year risk prediction horizon.

DISCUSSION

This is the first study to validate the comprehensive BOADICEA model¹² in predicting future BC risks using the joint effects of FH, QRFs, MD, PRS and rare PVs. The results show that the model is well calibrated both overall and in predicting the risks

Table 2 Calibration and discrimination of 5-year predicted breast cancer risks by menopausal status

Menopausal status	n. unaffected	n. BCs	Model	AUC (95% CI)	Harrell's C-index (95% CI)	E/O (95% CI)	Calibration slope (95% CI)
Subcohort with information on FH, QRFs, MD and PRS (n=15 502; n.BCs=676)							
Premenopausal	5502	212	FH+QRFs+PRS	0.68 (0.64 to 0.71)	0.67 (0.62 to 0.70)	1.10 (0.96 to 1.26)	1.02 (0.99 to 1.05)
			FH+QRFs+PRS+MD	0.69 (0.65 to 0.72)	0.68 (0.65 to 0.71)	1.14 (1.00 to 1.31)	1.03 (1.00 to 1.06)
Postmenopausal	9324	464	FH+QRFs+PRS	0.66 (0.63 to 0.68)	0.65 (0.63 to 0.68)	0.97 (0.89 to 1.06)	0.99 (0.97 to 1.02)
			FH+QRFs+PRS+MD	0.67 (0.64 to 0.69)	0.67 (0.64 to 0.69)	0.81 (0.74 to 0.89)	0.95 (0.93 to 0.97)
Subcohort with information on FH, QRFs, MD, PRS and PV status (n=5693; n.BCs=280)							
Premenopausal	1216	77	FH+QRFs+PRS+PV	0.70 (0.62 to 0.79)	0.68 (0.62 to 0.74)	1.11 (0.77 to 1.60)	1.02 (0.99 to 1.06)
			FH+QRFs+PRS+PV+MD	0.69 (0.62 to 0.76)	0.69 (0.62 to 0.76)	1.13 (0.79 to 1.62)	1.02 (0.99 to 1.05)
Postmenopausal	4197	203	FH+QRFs+PRS+PV	0.68 (0.63 to 0.73)	0.69 (0.66 to 0.72)	0.94 (0.79 to 1.13)	0.99 (0.96 to 1.01)
			FH+QRFs+PRS+PV+MD	0.68 (0.64 to 0.73)	0.69 (0.66 to 0.73)	0.78 (0.65 to 0.93)	0.94 (0.92 to 0.96)

BC, breast cancer; C-index, concordance index; E, expected number of BCs in the 5-year period; FH, family history; MD, mammographic density in BI-RADS; n.BCs, number of patients with incident breast cancer; n.unaffected, number of unaffected women; O, observed number BCs; PRS, polygenic risk score; PV, pathogenic variants in *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1*; QRFs, questionnaire-based risk factors.

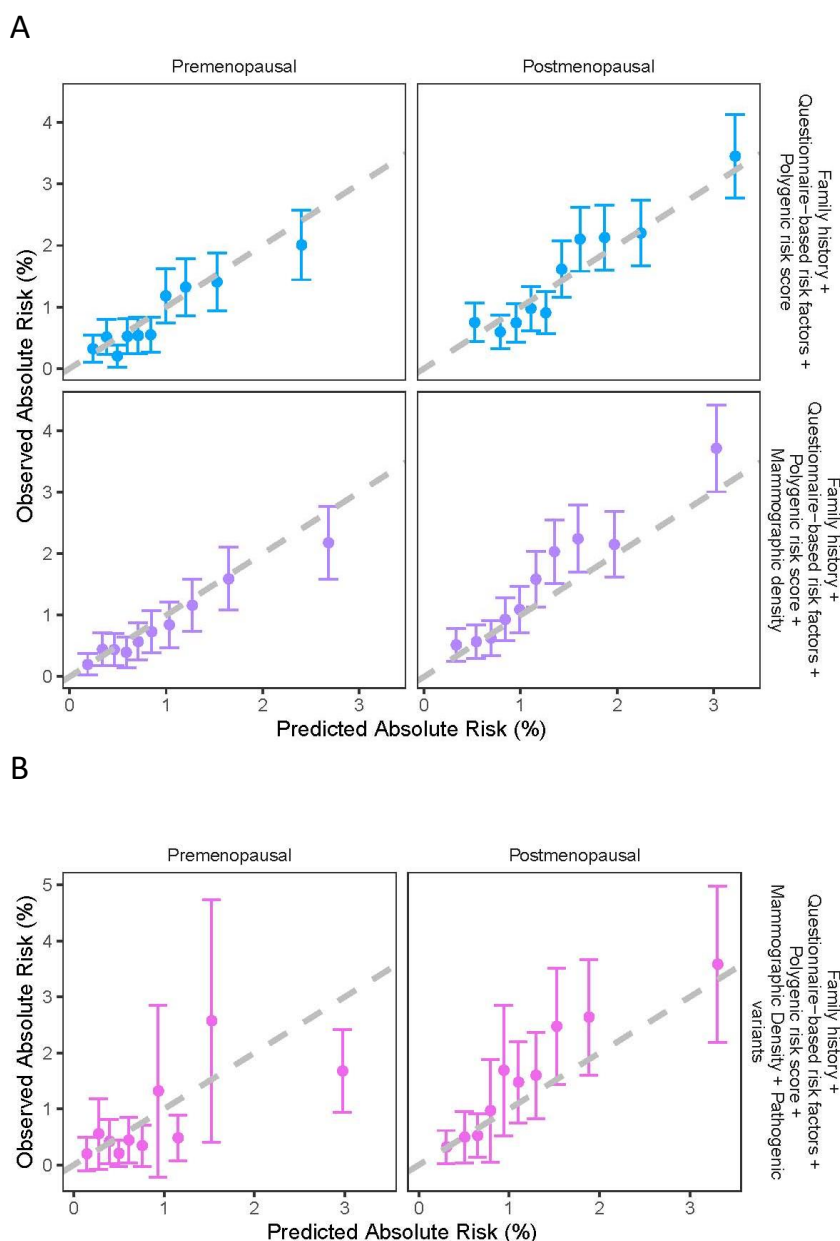


Figure 4 Observed and predicted 5-year breast cancer (BC) by menopausal status, using (A) the subcohort of participants with polygenic risk score (PRS) information under the model considering family history, questionnaire-based risk factors and PRS and the model considering family history, questionnaire-based risk factors, PRS and mammographic density in BI-RADS and (B) the subcohort of participants with PRS and pathogenic variant status under the model considering family history, questionnaire-based risk factors, PRS, mammographic density in BI-RADS and pathogenic variant status in the eight major BC susceptibility genes: *BRCA1*, *BRCA2*, *PALB2*, *CHECK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1*. Women were grouped into deciles of predicted risks.

for low-risk and high-risk women. The model is well calibrated when all, or subsets of risk factors are included. However, the discrimination is maximised when all risk factors are considered with an AUC of 0.70 (table 1). When assessing each risk factor individually, the PRS contributed most to the BC risk stratification.

Previous validation studies^{18 19} evaluated older versions of BOADICEA¹⁶ and have shown good discriminatory ability and calibration in women with BC FH.¹⁹ Detailed multigenerational pedigree data were included in those studies, as opposed to only first-degree relatives in this study which may be less informative. Two smaller population-based studies have assessed BOADICEA

V.5 using only FH, QRFs and PRS.^{20 21} These studies suggested that FH, QRFs and PRS jointly predicted risks well in high-risk and low-risk groups and had AUCs of 0.62 for predicting 10-year²⁰ and 0.65–0.70 for predicting 5-year risks.²¹ As different populations were used, the reported AUCs, which depend on the age distribution and other characteristics of the populations, are not directly comparable across studies. In the present study, the results show that compared with previous versions of BOADICEA that considered only FH information, the full multifactorial model improves discrimination (AUC increases from 0.61 to 0.70, table 1) and identifies 3.6% women as high risk (5-year BC risk $\geq 3\%$) compared with 0.8% when only FH is considered.

The inclusion of the 313-SNP PRS led to the greatest improvement in discriminatory ability, followed by MD, QRFs and first-degree FH. These patterns are in line with theoretical expectations under the BOADICEA model¹² and results from other studies^{34 38–40} and further support the inclusion of PRS in routine multifactorial risk assessment in clinical practice, for improving risk stratification.⁴¹ Here, we used the 313-SNP PRS.¹⁴ However, BOADICEA allows for alternative PRS to be used as long as the population distribution characteristics of the PRS are known.¹²

This study also suggests that BOADICEA is well calibrated overall and in deciles of predicted risks for all risk factors individually except for MD where there was some underprediction in the middle deciles, mainly among postmenopausal women (figure 4). This may represent some misspecification of the BI-RADS classification in the KARMA cohort. The BI-RADS classification was derived from continuous MD measurements generated using STRATUS²³ based on specific ‘%MD’ category thresholds. It does not correspond exactly to the visually assessed BI-RADS categories. For example, the underprediction in the seventh–eighth deciles of predicted risks using the model considering FH, QRFs, PRS and MD in postmenopausal women may be due to some misclassification between BI-RADS levels B and C when converting ‘%MD’ to BI-RADS (online supplemental figure S6). Nevertheless, BOADICEA was well calibrated in the highest and lowest BC risk groups for both premenopausal and postmenopausal women. Even for the visually assessed BI-RADS, previous studies showed that there is substantial variability in BI-RADS categorisation between radiologists.⁴² Since %MD has a stronger association with BC than categorised BI-RADS,⁴³ we would expect an improvement in BC risk prediction when incorporating continuous MD as risk factor in the model in the future.

Provided a model is calibrated in predicting risks in different risk categories, its clinical utility for prevention or early detection will depend on its ability to risk stratify the population by identifying groups of individuals in the population with sufficient differences in absolute risk to warrant the adoption of certain interventions. Therefore, the clinical utility of a model depends both on the discriminatory ability as measured by the AUC and the population risk of the disease.^{38 44} For a common disease, such as BC here, the results demonstrate that even modest increases in AUC can lead to a substantial increase in overall levels of risk stratification which are clinically meaningful. Taking the 5-year BC risk of 3% as a high-risk threshold, recommended for the risk-reducing treatments,^{8 34} under the full model, 3.6% women were classified as high risk, identifying 8.9% of the patients with incident BC during the 5-year risk prediction horizon. Alternatively, taking the 5-year BC risk of 1.67% as a ‘moderate-risk’ threshold, recommended for preventive risk reducing treatments in the USA,^{4 45} 16.9% women in the study were classified above this threshold, including 34.0% patients with incident BC. Taking the 5-year BC risk of 0.33% as a very-low-risk threshold (equivalent to a relative risk of 0.3 relative to the population risk at age 50), 88.9% women were above the threshold, identifying 96.6% patients with incident BC. The remaining 11.1% women identified as very-low-risk may opt for less intensive screening. In practice, optimal risk thresholds may need to be determined which consider the balance of benefits and harms from specific screening or early detection or prevention options.⁴⁶

This study has some limitations.^{31 32} Information on the PRS and PV status was available only for subsets of the participants. To maximise power, the subset with genetic information included the majority of incident BC cases and a random sample of unaffected women from the entire cohort. Direct model evaluation

using these subcohorts would have been susceptible to bias if the sampling scheme was ignored in the analysis. Here, we used a weighted cohort analytical approach^{29 34} that considers the probability of a genotyped individual being sampled from the entire cohort, which yields unbiased estimates for the model discriminatory ability and calibration. This is demonstrated in the results for the risk prediction models that could be evaluated both in the entire cohort and the weighted subcohorts: models that included QRFs, FH and MD, individually or combined (table 1). The model discriminatory ability and calibration estimates were almost identical in the entire cohort and the two subcohorts indicating no evidence of bias due to sampling.

Previous studies indicated that the PRS and other risk factors modify the BC risks for PV carriers and can lead to significantly different levels of risk stratification specifically for PV carriers.^{47–51} Our results showed that PV carriers had a wider BC risk distribution than non-carriers (online supplemental figure S5). However, we were not able to assess the detailed model performance separately for women with high-penetrance and moderate-penetrance PVs due the small number of PV carriers in the dataset (only 13 carriers developed BCs). Although at population level the increase in AUC when including gene-panel testing information on top of other risk factors and the PRS is modest, the change in personalised risk and risk stratification is large for those carrying PVs. As an example, among the 110 PV carriers, prior to the inclusion of gene-panel testing information, 74 had 5-year risks <1.67%, 30 had risks between 1.67% and 3% and 6 had risks >3%. After the inclusion of the gene-panel testing information, these numbers were 28, 25 and 57, respectively. The corresponding estimated E/O was improved from 0.31 (95% CI: 0.13 to 0.74) when no gene-panel testing information was considered to 0.99 (95% CI: 0.42 to 2.35) after including gene-panel testing information (online supplemental table S8).^{15 43 52} Although the subcohort with gene-panel testing information was also included in the study by Dorling *et al*,²⁷ this is unlikely to bias the results here because the penetrance parameters for PVs in most of the genes included in BOADICEA V.6 had been previously obtained from other studies.^{12 13 17} In BOADICEA V.6, only the BC relative risks for *BARD1* and *ATM* variants depend solely on BRIDGES data.¹² The BC relative risks for *RAD51C* and *RAD51D* PVs used in BOADICEA come from the meta-analysis of the Yang *et al*⁵³ and Dorling *et al*,²⁷ where the estimates were virtually identical.

KARMA is based on women attending mammographic screening, so the participants tend to be more highly educated and have a healthier lifestyle and may not be entirely representative of the general population.⁵⁴ Nevertheless, the results will be applicable to similar populations participating in screening programmes, one of the main settings in which risk stratification is used. KARMA included a small number of women who had mammography under 40 years old (0.4% of women), the minimum age entering the Swedish national screening programme. After excluding these women, the results remained similar (online supplemental table S9–S10, figure S7–S8) in terms of calibration and discrimination. Finally, the KARMA study participants were from Sweden, with the majority being of European ancestry. Therefore, further validation studies are required to assess the model validity in other ethnic subgroups and other countries with different BC incidences.

In summary, the comprehensive BOADICEA model is well calibrated and discriminates well between patients with incident BC and unaffected women undergoing mammographic screening. It is therefore a clinically valid model that can be used to identify high-risk individuals who may benefit most from enhanced

screening or other preventive risk-reducing treatments. It can also be used to identify low-risk individuals who are unlikely to benefit from prevention options associated with adverse effects. This validated model can be used as the basis for risk-based feasibility or acceptability studies or randomised controlled trials for determining the optimal risk thresholds, the uptake of different interventions and for health economics analyses that evaluate the balance between benefits and harms and associated healthcare costs from introducing multifactorial risk assessment, including the PRS, at different healthcare levels.⁷ BOADICEA is available via the CanRisk webtool www.canrisk.org⁶ that can be used by healthcare professionals in personalising risk assessment to facilitate shared decision-making with patients on lifestyle changes, prevention or screening options for managing BC risk.

Author affiliations

¹Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK

²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

³Department of Molecular Medicine, Université Laval and CHU de Québec-Université Laval Research Center, Quebec City, Quebec, Canada

⁴Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

⁵Division of Molecular Pathology, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

⁶Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

⁷Department of Oncology, Södersjukhuset, Stockholm, Sweden

Acknowledgements We would like to thank all the participants, clinicians and other healthcare professionals who have contributed to the KARMA cohort.

Contributors ACA and PH conceived the study design and supervised the study. XY and ACA carried out the statistical analyses. ME and GL were responsible for the data management. XY, ACA, PH, DFE and ME interpreted the results and drafted the initial manuscript. All other authors collected data, provided samples or were involved in the genetic testing process. All authors had full access to all the data in the study, read the manuscript, provided comments and approved the final manuscript version to submit for publication. ACA is responsible for the overall content as guarantor.

Funding This work was supported by grants from Cancer Research UK (C12292/A20861 and PPRPGM-Nov20/100002); the Märit and Hans Rausing's Initiative Against Breast Cancer; the European Union's Horizon 2020 research and innovation programme under grant agreement numbers 633784 (B-CAST) and 634935 (BRIDGES); the PERSPECTIVE I&I project, which the Government of Canada funds through Genome Canada (#13529) and the Canadian Institutes of Health Research (#155865), the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation; the CHU de Quebec Foundation and the Ontario Research Fund and the NIHR Cambridge Biomedical Research Centre.

Competing interests ACA, DFE and AL are named inventors of BOADICEA V.5 licensed by Cambridge Enterprise. All remaining authors have declared no conflicts of interest.

Patient consent for publication Not applicable.

Ethics approval The ethical board at Karolinska Institutet approved the study. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data are available under the Findability, Accessibility, Interoperability, Reproducibility (FAIR) principles, where all data necessary to reproduce the aims of the study are available from Karolinska Institutet through the MTA form available at <https://karmastudy.org/contact/data-access/>

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and

is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Xin Yang <http://orcid.org/0000-0003-0037-3790>

Andrew Lee <http://orcid.org/0000-0003-0677-0252>

Antonis C Antoniou <http://orcid.org/0000-0001-9223-3116>

REFERENCES

- Breast Cancer - World Health Organization. Available: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>
- Francis PA, Regan MM, Fleming GF. Adjuvant ovarian suppression in premenopausal breast cancer. *N Engl J Med* 2015;372:1673-436-46.
- Mandrik O, Zielonke N, Meheus F, Severens JH, Guha N, Herrero Acosta R, Murillo R. Systematic reviews as a 'lens of evidence': Determinants of benefits and harms of breast cancer screening. *Int J Cancer* 2019;145:994-1006.
- Visvanathan K, Hurley P, Bantug E, Brown P, Col NF, Cuzick J, Davidson NE, Decensi A, Fabian C, Ford L, Garber J, Katapodi M, Kramer B, Morrow M, Parker B, Runowicz C, Vogel VG, Wade JL, Lippman SM. Use of pharmacologic interventions for breast cancer risk reduction: American Society of clinical oncology clinical practice guideline. *J Clin Oncol* 2013;31:2942-62.
- Pashayan N, Antoniou AC, Ivanus U, Esserman LJ, Easton DF, French D, Sroczynski G, Hall P, Cuzick J, Evans DG, Simard J, Garcia-Closas M, Schmutzler R, Wegwarth O, Pharoah P, Moorhith S, De Montgolfier S, Baron C, Hecceg Z, Turnbull C, Balleyguier C, Rossi PG, Wesseling J, Ritchie D, Tischkowitz M, Broeders M, Reis D, Metspalu A, Callender T, de Koning H, Devilee P, Delaloge S, Schmidt MK, Widschwendter M. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol* 2020;17:687-705.
- Carver T, Hartley S, Lee A, Cunningham AP, Archer S, Babb de Villiers C, Roberts J, Ruston R, Walter FM, Tischkowitz M, Easton DF, Antoniou AC. CanRisk Tool-A web interface for the prediction of breast and ovarian cancer risk and the likelihood of carrying genetic pathogenic variants. *Cancer Epidemiol Biomarkers Prev* 2021;30:469-73.
- Brooks JD, Nabi HH, Andrusis IL, Antoniou AC, Chiquette J, Després P, Devilee P, Dorval M, Droit A, Easton DF, Eisen A, Eloy L, Fienberg S, Goldgar D, Hahnen E, Joly Y, Knoppers BM, Lofers A, Masson J-Y, Mittmann N, Paquette J-S, Pashayan N, Schmutzler R, Stockley T, Tavtigian SV, Walker MJ, Wolfson M, Chiarelli AM, Simard J. Personalized Risk Assessment for Prevention and Early Detection of Breast Cancer: Integration and Implementation (PERSPECTIVE I&I). *J Pers Med* 2021;11. doi:10.3390/jpm11060511. [Epub ahead of print: 04 Jun 2021].
- The National Institute for Health and Care Excellence (NICE). Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer, 2013. Available: <https://www.nice.org.uk/guidance/cg164>
- Public Health England. *Protocols for surveillance of women at very high risk of developing breast cancer*, 2022. <https://www.gov.uk/government/publications/breast-screening-higher-risk-women-surveillance-protocols/protocols-for-surveillance-of-women-at-higher-risk-of-developing-breast-cancer>
- National Comprehensive Cancer Network. *NCCN Guidelines - Breast Cancer Risk Reduction*, 2022. <https://www.nccn.org/guidelines/guidelines-detail?category=2&id=1420>
- Cancer Care Ontario. *Breast cancer screening for women at high risk*, 2022. <https://www.cancercareontario.ca/en/guidelines-advice/cancer-continuum/screening/breast-cancer-high-risk-women>
- Lee A, Mavaddat N, Cunningham AP, Carver T, Archer S, Walter FM, Tischkowitz M, Roberts J, Usher-Smith J, Simard J, Schmidt MK, Devilee P, Zadnik V, Jürgens H, Mouret-Fourme E, De Pauw A, Rookus M, Mooij TM, Pharoah PPD, Easton DF, Antoniou AC. Enhancing the BOADICEA cancer risk prediction model to incorporate new data on *RAD51C*, *RAD51D*, *BARD1* updates to tumour pathology and cancer incidences. *J Med Genet* 2022;59:1205-17.
- Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, Babb de Villiers C, Izquierdo A, Simard J, Schmidt MK, Walter FM, Chatterjee N, Garcia-Closas M, Tischkowitz M, Pharoah P, Easton DF, Antoniou AC. Boadicea: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med* 2019;21:1708-18. doi:10.1038/s41436-018-0406-9
- Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen T-H, Wang Q, Bolla MK, Yang X, Adank MA, Ahearn T, Aittomäki K, Allen J, Andrulis IL, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Auer PL, Auvinen P, Barndahl M, Beane Freeman LE, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Borresen-Dale A-L, Brauch H, Bremer M, Brenner H, Brentnall A, Brock IW, Brooks-Wilson A, Brucker SY,

- Brüning T, Burwinkel B, Campa D, Carter BD, Castelao JE, Chanock SJ, Chlebowski R, Christiansen H, Clarke CL, Collée JM, Cordina-Duverger E, Cornelissen S, Couch FJ, Cox A, Cross SS, Czene K, Daly MB, Devilee P, Dörk T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Engel C, Eriksson M, Evans DG, Fasching PA, Figueroa J, Fletcher O, Flyger H, Försti A, Fritschl L, Gabrielson M, Gago-Dominguez M, Gapstur SM, García-Sáenz JA, Gaudet MM, Georgoulas V, Giles GG, Gilyazova IR, Glendon G, Goldberg MS, Goldgar DE, González-Neira A, Grenaker Alnæs GI, Grip M, Gronwald J, Grundy A, Guénel P, Haebeler L, Hahnen E, Haiman CA, Håkansson N, Hamann U, Hankinson SE, Harkness EF, Hart SN, He W, Hein A, Heyworth J, Hillemanns P, Hollestelle A, Hoening MJ, Hoover RN, Hopper JL, Howell A, Huang G, Humphreys K, Hunter DJ, Jakimovska M, Jakubowska A, Janni W, John EM, Johnson N, Jones ME, Jukkola-Vuorinen A, Jung A, Kaaks R, Kaczmarsk K, Kataja V, Keeman R, Kerin MJ, Khushnudinova E, Kiiski JI, Knight JA, Ko Y-D, Kosma V-M, Koutros S, Kristensen VN, Krüger U, Kühn T, Lambrechts D, Le Marchand L, Lee E, Lejbkowitz F, Lilyquist J, Lindblom A, Lindström S, Lissowska J, Lo W-Y, Lubiński J, Long J, Lubiński J, Lux MP, MacInnis RJ, Maishman T, Makalic E, Maleva Kostovska I, Mannermaa A, Manoukian S, Margolin S, Martens JWM, Martinez ME, Mavroudis D, McLean C, Meindl A, Menon U, Middha P, Miller N, Moreno F, Mulligan AM, Mulot C, Muñoz-Garzon VM, Neuhausen DL, Nevanlinna H, Neven P, Newman WG, Nielsen SF, Nordestgaard BG, Norman A, Offit K, Olson JE, Olsson H, Orr N, Pankratz VS, Park-Simon T-W, Perez JIA, Pérez-Barrios C, Peterlongo P, Peto J, Pinchev M, Plaseska-Karanfilska D, Polley EC, Prentice R, Presneau N, Prokofyeva D, Purrington K, Pykäs K, Rack B, Radice P, Rau-Murthy R, Rennett HS, Rhenius V, Robson M, Romero A, Ruddy KJ, Ruebner M, Saloustros E, Sandler DP, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schürmann P, Schwentner L, Scott C, Scott RJ, Seynaeve C, Shah M, Sherman ME, Shrubsole MJ, Shu X-O, Slager S, Smeets A, Sohn C, Soucy P, Southey MC, Spinelli JJ, Stegmaier C, Stone J, Swerdlow AJ, Tamimi RM, Tapper WJ, Taylor JA, Terry MB, Thöne K, Tollenaar RAEM, Tomlinson I, Truong T, Tzardi M, Ulmer H-U, Untch M, Vachon CM, van Veen EM, Vijai J, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Yang XR, Yannoukakos D, Zhang Y, Zheng W, Zogas A, Dunning AM, Thompson DJ, Chenevix-Trench G, Chang-Claude J, Schmidt MK, Hall P, Milne RL, Pharoah PDP, Antoniou AC, Chatterjee N, Kraft P, Garcia-Closas M, Simard J, Easton DF, ABCB Investigators, kConFab/AOCS Investigators, NBCS Collaborators. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet* 2019;104:21–34.
- 15 Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. BI-RADS® fifth edition: A summary of changes. *Diagn Interv Imaging* 2017;98:179–90.
- 16 Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA, Risch HA, Eyfjord JE, Hopper JL, Southey MC, Olsson H, Johansson O, Borg A, Pasini B, Passini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tryggvadottir L, Syrjäkoski K, Kallioniemi O-P, Eerola H, Nevanlinna H, Pharoah PDP, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* 2008;98:1457–66.
- 17 Lee AJ, Cunningham AP, Tischkowitz M, Simard J, Pharoah PD, Easton DF, Antoniou AC. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. *Genet Med* 2016;18:1190–8.
- 18 MacInnis RJ, Bickerstaffe A, Apicella C, Dite GS, Dowty JG, Aujard K, Phillips K-A, Weideman P, Lee A, Terry MB, Giles GG, Southey MC, Antoniou AC, Hopper JL. Prospective validation of the breast cancer risk prediction model BOADICEA and a batch-mode version BOADICEACentre. *Br J Cancer* 2013;109:1296–301.
- 19 Terry MB, Liao Y, Whittemore AS, Leocne N, Buchsbaum R, Zeinomar N, Dite GS, Chung WK, Knight JA, Southey MC, Milne RL, Goldgar D, Giles GG, McLachlan S-A, Friedlander ML, Weideman PC, Glendon G, Nesci S, Andrulis IL, John EM, Phillips K-A, Daly MB, Buys SS, Hopper JL, MacInnis RJ. 10-Year performance of four models of breast cancer risk: a validation study. *Lancet Oncol* 2019;20:504–17.
- 20 Li SX, Milne RL, Nguyen-Dumont T, Wang X, English DR, Giles GG, Southey MC, Antoniou AC, Lee A, Li S, Winship I, Hopper JL, Terry MB, MacInnis RJ. Prospective evaluation of the addition of polygenic risk scores to breast cancer risk models. *JNCI Cancer Spectr* 2021;5. doi:10.1093/jncics/pkab021. [Epub ahead of print: 02 Mar 2021].
- 21 Pal Choudhury P, Brook MN, Hurson AN, Lee A, Mulder CV, Coulson P, Schoemaker MJ, Jones ME, Swerdlow AJ, Chatterjee N, Antoniou AC, Garcia-Closas M. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. *Breast Cancer Res* 2021;23.
- 22 Gabrielson M, Eriksson M, Hammarström 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;46:1740–1.
- 23 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;169:371–9.
- 24 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.
- 25 Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med* 2008;148:337–47.
- 26 Lofgren L, Eloranta S, Krawiec K, Asterkvist A, Lonnqvist C, Sandelin K. Steering group of the National register for breast C. validation of data quality in the Swedish national register for breast cancer. *BMC Public Health* 2019;19:495.
- 27 Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, Pooley KA, Parsons MT, Fortunato C, Wang Q, Bolla MK, Dennis J, Keeman R, Alonso MR, Álvarez N, Herraez B, Fernandez V, Núñez-Torres R, Osorio A, Valcich J, Li M, Törngren T, Harrington PA, Baynes C, Conroy DM, Decker B, Fachal L, Mavaddat N, Ahearn T, Aittomäki K, Anttonenkovva NN, Arnold N, Arveux P, Ausems MGEM, Auvinen P, Becher H, Beckmann MW, Behrens S, Bermisheva M, Bialkowska K, Blomqvist C, Bogdanova NV, Bogdanova-Markov N, Bojesen SE, Bonanni B, Borresen-Dale A-L, Brauch H, Bremer M, Briceno I, Brüning T, Burwinkel B, Cameron DA, Camp NJ, Campbell A, Carracedo A, Castelao JE, Cessna MH, Chanock SJ, Christiansen H, Collée JM, Cordina-Duverger E, Cornelissen S, Czene K, Dörk T, Ekici AB, Engel C, Eriksson M, Fasching PA, Figueroa J, Flyger H, Försti A, Gabrielson M, Gago-Dominguez M, Georgoulas V, Gil F, Giles GG, Glendon G, Garcia EBG, Alnæs GI, Guénel P, Hadjisavvas A, Haebeler L, Hahnen E, Hall P, Hamann U, Harkness EF, Hartikainen JM, Hartman M, He W, Heemskerk-Gerritsen BAM, Hillemanns P, Hogervorst FBL, Hollestelle A, Ho WK, Hoening MJ, Howell A, Humphreys K, Idris F, Jakubowska A, Jung A, Kapoor PM, Kerin MJ, Khushnudinova E, Kim S-K, Kosma V-M, Kristensen VN, Kyriacou K, Lakeman IMM, Lee JW, Lee MH, Li J, Lindblom A, Lo W-Y, Loizidou MA, Lophatananon A, Lubiński J, MacInnis RJ, Madsen MJ, Mannermaa A, Manoochehri M, Manoukian S, Margolin S, Martinez ME, Maurer T, Mavroudis D, McLean C, Meindl A, Mensenkamp AR, Michailidou K, Miller N, Mohd Taib NA, Muir K, Mulligan AM, Nevanlinna H, Newman WG, Nordestgaard BG, Ng P-S, Oosterwijk JC, Park SK, Park-Simon T-W, Perez JIA, Peterlongo P, Porteous DJ, Prajezandanc K, Prokofyeva D, Radice P, Rashid MU, Rhenius V, Rookus MA, Rüdigier T, Saloustros E, Sawyer EJ, Schmutzler RK, Schneeweiss A, Schürmann P, Shah M, Sohn C, Southey MC, Surowy H, Suvanto M, Thanassitichai S, Tomlinson I, Torres D, Truong T, Tzardi M, Valova Y, van Asperen CJ, Van Dam RM, van den Ouweland AMW, van der Kolk LE, van Veen EM, Wendt C, Williams JA, Yang XR, Yoon S-Y, Zamora MP, Evans DG, de la Hoya M, Simard J, Antoniou AC, Borg Åke, Andrulis IL, Chang-Claude J, Garcia-Closas M, Chenevix-Trench G, Milne RL, Pharoah PDP, Schmidt MK, Spurdle AB, Vreeswijk MPG, Benitez J, Dunning AM, Kvist A, Teo SH, Devilee P, Easton DF. Breast Cancer Association Consortium. Breast Cancer Risk Genes – Association Analysis in More than 113,000 Women. *N Engl J Med* 2021;384:428–39.
- 28 Hayward NK, Wilmott JS, Waddell N, Johansson PA, Field MA, Nones K, Patch A-M, Kakavand H, Alexandrov LB, Burke H, Jakrot V, Kazakoff S, Holmes O, Leonard C, Sabarinathan R, Mularoni L, Wood S, Xu Q, Waddell N, Tembe V, Pupo GM, De Paoli-Iseppi R, Vilain RE, Shang P, Lau LMS, Dagg RA, Schramm S-J, Pritchard A, Dutton-Regester K, Newell F, Fitzgerald A, Shang CA, Grimmond SM, Pickett HA, Yang JY, Stretch JR, Behren A, Kefford RF, Hersey P, Long GV, Cebon J, Shackleton M, Spillane AJ, Saw RPM, López-Bigas N, Pearson JV, Thompson JF, Scolyer RA, Mann GJ. Whole-Genome landscapes of major melanoma subtypes. *Nature* 2017;545:175–80.
- 29 Pal Choudhury P, Maas P, Wilcox A, Wheeler W, Brook M, Check D, Garcia-Closas M, Chatterjee N. iCARE: an R package to build, validate and apply absolute risk models. *PLoS One* 2020;15:e0228198.
- 30 Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 2001;93:358–66.
- 31 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- 32 Pencina MJ, D'Agostino RB, Song L. Quantifying discrimination of Framingham risk functions with different survival C statistics. *Stat Med* 2012;31:1543–53.
- 33 White IR, Rapsonnik E. Emerging risk factors C. Covariate-adjusted measures of discrimination for survival data. *Biom J* 2015;57:592–613.
- 34 Hurson AN, Pal Choudhury P, Gao C, Hüsing A, Eriksson M, Shi M, Jones ME, Evans DGR, Milne RL, Gaudet MM, Vachon CM, Chasman DI, Easton DF, Schmidt MK, Kraft P, Garcia-Closas M, Chatterjee N, B-CAST Risk Modelling Group. Prospective evaluation of a breast-cancer risk model integrating classical risk factors and polygenic risk in 15 cohorts from six countries. *Int J Epidemiol* 2022;50:1897–911.
- 35 Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, Caughey AB, Doubeni CA, Epling JW, Kubik M, Landefeld CS, Mangione CM, Pbert L, Silverstein M, Tseng C-W, Wong JB, US Preventive Services Task Force. Medication use to reduce risk of breast cancer: US preventive services Task force recommendation statement. *JAMA* 2019;322:857–67.
- 36 R Core Team. R: A Language and Environment for Statistical Computing. In: Vienna A, ed. *Computing RFS*, 2020. <https://www.R-project.org/>
- 37 Stevenson M, ESwcft N, Heuer C, Marshall J, Sanchez J, Thornton R, Reiczgel J, Robison-Cox J, Sebastiani P, Solymos P, Yoshida K, Jones G, Pirikahu S, Firestone S, Kyle R, Popp J, Jay M, Reynard C. *epiR: tools for the analysis of epidemiological data*, 2021. <https://CRAN.R-project.org/package=epiR>
- 38 Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst* 2014;106. doi:10.1093/jnci/dju305. [Epub ahead of print: 12 Nov 2014].
- 39 Maas P, Barrdahl M, Joshi AD, Auer PL, Gaudet MM, Milne RL, Schumacher FR, Anderson WF, Check D, Chattopadhyay S, Baglietto L, Berg CD, Chanock SJ, Cox DG, Figueroa JD, Gail MH, Graubard BI, Haiman CA, Hankinson SE, Hoover RN, Isaacs

- C, Kolonel LN, Le Marchand L, Lee I-M, Lindström S, Overvad K, Romieu I, Sanchez M-J, Southey MC, Stram DO, Tumino R, VanderWeele TJ, Willett WC, Zhang S, Buring JE, Canzian F, Gapstur SM, Henderson BE, Hunter DJ, Giles GG, Prentice RL, Ziegler RG, Kraft P, Garcia-Closas M, Chatterjee N. Breast cancer risk from modifiable and Nonmodifiable risk factors among white women in the United States. *JAMA Oncol* 2016;2:1295–302.
- 40 Kachuri L, Graff RE, Smith-Byrne K, Meyers TJ, Rashkin SR, Ziv E, Witte JS, Johansson M. Pan-Cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun* 2020;11:6084.
- 41 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 2018;19:581–90.
- 42 Gard CC, Aiello Bowles EJ, Miglioretti DL, Taplin SH, Rutter CM. Misclassification of breast imaging reporting and data system (BI-RADS) mammographic density and implications for breast density reporting legislation. *Breast J* 2015;21:481–9.
- 43 McCormack VA, dos Santos Silva I, Silva dos S I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006;15:1159–69.
- 44 Clayton DG. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS Genet* 2009;5:e1000540.
- 45 Freedman AN, Graubard BI, Rao SR, McCaskill-Stevens W, Ballard-Barbash R, Gail MH. Estimates of the number of US women who could benefit from tamoxifen for breast cancer chemoprevention. *J Natl Cancer Inst* 2003;95:526–32.
- 46 Fitzgerald RC, Antoniou AC, Fruk L, Rosenfeld N. The future of early cancer detection. *Nat Med* 2022;28:666–77.
- 47 Muranen TA, Greco D, Blomqvist C, Aittomäki K, Khan S, Hogervorst F, Verhoef S, Pharoah PDP, Dunning AM, Shah M, Luben R, Bojesen SE, Nordestgaard BG, Schoemaker M, Swerdlow A, Garcia-Closas M, Figueroa J, Dörk T, Bogdanova NV, Hall P, Li J, Khusnutdinova E, Bermisheva M, Kristensen V, Borresen-Dale A-L, Peto J, Dos Santos Silva I, Couch FJ, Olson JE, Hillemans P, Park-Simon T-W, Brauch H, Hamann U, Burwinkel B, Marme F, Meindl A, Schmutzler RK, Cox A, Cross SS, Sawyer EJ, Tomlinson I, Lambrechts D, Moisse M, Lindblom A, Margolin S, Hollestelle A, Martens JWM, Fasching PA, Beckmann MW, Andrulis IL, Knight JA, Anton-Culver H, Ziogas A, Giles GG, Milne RL, Brenner H, Arndt V, Mannermaa A, Kosma V-M, Chang-Claude J, Rudolph A, Devilee P, Seynaeve C, Hopper JL, Southey MC, John EM, Whittemore AS, Bolla MK, Wang Q, Michailidou K, Dennis J, Easton DF, Schmidt MK, Nevanlinna H, NBCS Investigators, kConFab/AOCS Investigators. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. *Genet Med* 2017;19:599–603.
- 48 Barnes DR, Rookus MA, McGuffog L, Leslie G, Mooij TM, Dennis J, Mavaddat N, Adlard J, Ahmed M, Aittomäki K, Andrieu N, Andrulis IL, Arnold N, Arun BK, Azzollini J, Balmana J, Barkardottir RB, Barrowdale D, Benitez J, Berthet P, Bialkowska K, Blanco AM, Blok MJ, Bonanni B, Boonen SE, Borg A, Bozsik A, Bradbury AR, Brennan P, Brewer C, Brunet J, Buys SS, Caldes T, Caligo MA, Campbell I, Christensen LL, Chung WK, Claes KBM, Colas C, Collaborators GS, Collaborators E, Collonge-Rame MA, Cook J, Daly MB, Davidson R, de la Hoya M, de Putter R, Delnatte C, Devilee P, Diez O, Ding YC, Domchek SM, Dorfling CM, Dumont M, Eeles R, Ejlersen B, Engel C, Evans DG, Faivre L, Foretova L, Fostira F, Friedlander M, Friedman E, Frost D, Ganz PA, Garber J, Gehrig A, Gerdes AM, Gesta P, Giraud S, Glendon G, Godwin AK, Goldgar DE, Gonzalez-Neira A, Greene MH, Gschwanter-Kaulich D, Hahnen E, Hamann U, Hanson H, Hentschel J, Hogervorst FBL, Hoening MJ, Horvath J, Hu C, Hulick PJ, Ilyanov EN, kConFab I, Investigators H, Investigators G, Isaacs C, Izatt L, Izquierdo A, Jakubowska A, James PA, Janavicius R, John EM, Joseph V, Karlan BY, Kast K, Koudijs M, Kruse TA, Kwong A, Laitman Y, Lasset C, Lazaro C, Lester J, Lesueur F, Liljegren A, Loud JT, Lubinski J, Mai PL, Manoukian S, Mari V, Mebirouk N, Meijers-Heijboer HEJ, Meindl A, Mensenkamp AR, Miller A, Montagna M, Mouret-Fourme E, Mukherjee S, Mulligan AM, Nathanson KL, Neuhausen SL, Nevanlinna H, Niederacher D, Nielsen FC, Nikitina-Zake L, Nogues C, Olah E, Olopade OI, Ong KR, O'Shaughnessy-Kirwan A, Osorio A, Ott CE, Papi L, Park SK, Parsons MT, Pedersen IS, Peissel B, Peixoto A, Peterlongo P, Pfeiler G, Phillips KA, Prajzandanc K, Pujana MA, Radice P, Ramser J, Ramus SJ, Rantala J, Rennett G, Risch HA, Robson M, Ronlund K, Salani R, Schuster H, Senter L, Shah PD, Sharma P, Side LE, Singer CF, Slavin TP, Soucy P, Southey MC, Spurdle AB, Steinemann D, Steinsnyder Z, Stoppa-Lyonnet D, Sutter C, Tan YY, Teixeira MR, Teo SH, Thull DL, Tischkowitz M, Tognazzo S, Toland AE, Trainer AH, Tung N, van Engelen K, van Rensburg EJ, Vega A, Vierstraete J, Wagner G, Walker L, Wang-Gohrke S, Wappenschmidt B, Weitzel JN, Yadav S, Yang X, Yannoukakos D, Zimbalatti D, Offit K, Thomassen M, Couch FJ, Schmutzler RK, Simard J, Easton DF, Chenevix-Trench G, Antoniou AC. Consortium of Investigators of modifiers of B, BRCA. polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet Med* 2020;22:1653–66.
- 49 Andrieu N, Goldgar DE, Easton DF, Rookus M, Brohet R, Antoniou AC, Peock S, Evans G, Eccles D, Douglas F, Noguès C, Gauthier-Villars M, Chompret A, Van Leeuwen FE, Kluijdt I, Benitez J, Arver B, Olah E, Chang-Claude J, Wagner G, Geo H, Pregnancies GIC. EMBRACE, GENEPSO, GEO-HEBON, IBCCS Collaborators Group. Pregnancies, breast-feeding, and breast cancer risk in the International BRCA1/2 carrier cohort study (IBCCS). *J Natl Cancer Inst* 2006;98:535–44.
- 50 Brohet RM, Goldgar DE, Easton DF, Antoniou AC, Andrieu N, Chang-Claude J, Peock S, Eeles RA, Cook M, Chu C, Noguès C, Lasset C, Berthet P, Meijers-Heijboer H, Gerdes A-M, Olsson H, Caldes T, van Leeuwen FE, Rookus MA. Oral contraceptives and breast cancer risk in the International BRCA1/2 carrier cohort study: a report from embrace, GENEPSO, GEO-HEBON, and the IBCCS collaborating group. *J Clin Oncol* 2007;25:3831–6.
- 51 Chang-Claude J, Andrieu N, Rookus M, Brohet R, Antoniou AC, Peock S, Davidson R, Izatt L, Cole T, Noguès C, Luporsi E, Huiart L, Hoogerbrugge N, Van Leeuwen FE, Osorio A, Eyfjord J, Radice P, Goldgar DE, Easton DF. Epidemiological Study of Familial Breast Cancer (EMBRACE), Gene Etude Prospective Sein Ovaire (GENEPSO), Genen Omgeving studie van de werkgroep Hereditair Borstkanker Onderzoek Nederland (GEO-HEBON), International BRCA1/2 Carrier Cohort Study (IBCCS) collaborators group. Age at menarche and menopause and breast cancer risk in the International BRCA1/2 carrier cohort study. *Cancer Epidemiol Biomarkers Prev* 2007;16:740–6.
- 52 Eriksson M, Czene K, Strand F, Zackrisson S, Lindholm P, Lång K, Förnvik D, Sartor H, Mavaddat N, Easton D, Hall P. Identification of women at high risk of breast cancer who need supplemental screening. *Radiology* 2020;297:327–33.
- 53 Yang X, Song H, Leslie G, Engel C, Hahnen E, Auber B, Horváth J, Kast K, Niederacher D, Turnbull C, Houlston R, Hanson H, Loveday C, Dolinsky JS, LaDuca H, Ramus SJ, Menon U, Rosenthal AN, Jacobs I, Gayther SA, Dicks E, Nevanlinna H, Aittomäki K, Pelttari LM, Ehrencrona H, Borg Åke, Kvist A, Rivera B, Hansen TVO, Djursby M, Lee A, Dennis J, Bowtell DD, Traficante N, Diez O, Balmaña J, Gruber SB, Chenevix-Trench G, Investigators kConFab, Jensen A, Kjær SK, Høgdall E, Castéra L, Garber J, Janavicius R, Osorio A, Golmard L, Vega A, Couch FJ, Robson M, Gronwald J, Domchek SM, Culver JO, de la Hoya M, Easton DF, Foulkes WD, Tischkowitz M, Meindl A, Schmutzler RK, Pharoah PDP, Antoniou AC. Ovarian and breast cancer risks associated with pathogenic variants in RAD51C and RAD51D. *J Natl Cancer Inst* 2020;112:1242–50.
- 54 Azam S, Eriksson M, Sjölander A, Hellgren R, Gabrielson M, Czene K, Hall P. Mammographic density change and risk of breast cancer. *J Natl Cancer Inst* 2020;112:391–9.

Supplementary Materials

METHODS

SNP genotyping and gene-panel testing

Blood samples were collected at recruitment and were processed in the Karolinska Institutet high-throughput Biobank[1]. A subset of the samples, including the majority of BCs, were genotyped using either iCOGS[2, 3] or OncoArray[4] in studies organised through the Breast Cancer Association Consortium (BCAC). Here, we used only the prospective arm of the KARMA study. Genotype quality control was performed as previously described[5]. Genotypes for SNPs not presented on the arrays were imputed using the 1000 Genomes Project (phase 3) as the reference panel[5]. The 313-SNP PRS was derived and standardised by the mean (-0.424) and standard deviation (0.603) as described by Mavaddat et al.[6]. A subset of these samples was sequenced for a 34-gene panel through the Breast Cancer Risk after Diagnostic Gene Sequencing (BRIDGES) study. Details of library preparation, sample sequencing, variant calling and quality control have been described previously[7]. Women were considered as PV carriers in *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *RAD51C*, *RAD51D*, or *BARD1* if a protein-truncating (frameshift, nonsense or canonical splice site variant) or a known pathogenic missense variant was identified, with the exception of variants identified in the last exon of each gene.

REFERENCES

1. Gabrielson M, Eriksson M, Hammarstrom M, et al. Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). *Int J Epidemiol* 2017;**46**(6):1740-41g doi: 10.1093/ije/dyw357[published Online First: 2017/02/10].
2. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015;**47**(4):373-80 doi: 10.1038/ng.3242[published Online First: 2015/03/10].
3. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;**45**(4):353-61, 61e1-2 doi: 10.1038/ng.2563[published Online First: 2013/03/29].
4. Amos CI, Dennis J, Wang Z, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev* 2017;**26**(1):126-35 doi: 10.1158/1055-9965.EPI-16-0106
5. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015;**526**(7571):68-74 doi: 10.1038/nature15393[published Online First: 2015/10/04].

6. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2019;**104**(1):21-34 doi: 10.1016/j.ajhg.2018.11.002[published Online First: 2018/12/18].
7. Breast Cancer Association C, Dorling L, Carvalho S, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* 2021;**384**(5):428-39 doi: 10.1056/NEJMoa1913948[published Online First: 2021/01/21].
8. Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med* 2008;**148**(5):337-47 doi: 10.7326/0003-4819-148-5-200803040-00004[published Online First: 2008/03/05].

Table s1 Comparison of distribution of STRATUS dereived BI-RADS categories in the KARMA cohort and previously published study[8].

BI-RADS Category	In KARMA	Tice <i>et al.</i> [8]
Age < 50y		
A	0.05	0.04
B	0.24	0.35
C	0.50	0.47
D	0.20	0.14
Age 50-64y		
A	0.13	0.07
B	0.42	0.47
C	0.38	0.40
D	0.06	0.06
Age ≥ 65y		
A	0.21	0.11
B	0.50	0.54
C	0.26	0.32
D	0.03	0.03

Table s2 A summary of genetic and epidemiological characteristics of participants at baseline in (1) the entire KARMA cohort (2) sub-cohort samples with PRS information and (3) sub-cohort samples with both PRS and pathogenic variant (PV) status in *BRCA1*, *BRCA2*, *PALB2*, *CHECK2*, *ATM*, *RAD51C*, *RAD51D*, and *BARD1*. The incident BC patients were women who developed breast cancer within five-year risk prediction horizon.

	Full cohort		Sub-cohort with PRS		Sub-cohort with PRS and PV	
	Healthy women	Incident BC patients	Healthy women	Incident BC patients	Healthy women	Incident BC patients
Number of participants, <i>N</i>						
	65,599	816	14,826	676	5,413	280
Follow-up, median (IQR), years						
	6.0 (5.0-6.0)	3.0 (2.0-4.0)	6.0 (5.0-6.0)	3.0 (2.0-4.0)	6.0 (5.0-6.0)	3.0 (2.0-4.0)
European ancestry, <i>N</i> (%)						
Yes	59,102 (90.1%)	757 (92.8%)	14,264 (96.2%)	633 (93.6%)	5,307 (98.0%)	267 (95.4%)
No	1,642 (2.5%)	16 (2.0%)	223 (1.5%)	9 (1.3%)	2 (0.04%)	4 (1.4%)
Missing	4,855 (7.4%)	43 (5.2%)	339 (2.3%)	34 (5.1%)	104 (1.9%)	9 (3.2%)
Age at baseline, <i>N</i> (%)						
<40	263 (0.4%)	0 (0%)	75 (0.5%)	0 (0%)	40 (0.7%)	0 (0.0%)
40–49	23,618 (36.0%)	177 (21.7%)	4,048 (27.3%)	140 (20.7%)	695 (12.8%)	56 (20.0%)
50–59	18,275 (27.9%)	231 (28.3%)	3,951 (26.7%)	196 (29.0%)	1,338 (24.7%)	77 (27.5%)
≥60	23,443 (35.7%)	408 (50.0%)	6,752 (45.5%)	340 (50.3%)	3,340 (61.7%)	147 (52.5%)
Mean (sd), years	55 (9.8)	58 (9.4)	57 (10.0)	59 (9.3)	61 (9.1)	59 (9.2)
Age at menarche*, <i>N</i> (%)						
<11	6,828 (10.4%)	79 (9.7%)	728 (4.9%)	62 (9.2%)	250 (4.6%)	18 (6.4%)
[11,12)	6,251 (9.5%)	69 (8.5%)	1,487 (10.0%)	53 (7.8%)	512 (9.5%)	27 (9.6%)
[12,13)	13,085 (20.0%)	164 (20.1%)	3,106 (21.0%)	132 (19.5%)	1,043 (19.3%)	50 (17.9%)
[13,14)	16,759 (25.5%)	216 (26.5%)	3,935 (26.5%)	189 (28.0%)	1,403 (25.9%)	77 (27.5%)
[14,15]	13,127 (20.0%)	175 (21.4%)	3,210 (21.6%)	148 (21.9%)	1,230 (22.7%)	70 (25.0%)

	[15,16]	6,625 (10.1%)	74 (9.1%)	1,654 (11.2%)	61 (9.0%)	692 (12.8%)	27 (9.6%)
	≥16	2,924 (4.5%)	39 (4.8%)	706 (4.8%)	31 (4.6%)	283 (5.2%)	11 (3.9%)
Menopausal status, N (%)							
	Pre-menopausal	29,829 (45.5%)	268 (32.8%)	5,502 (37.1%)	212 (31.4%)	1,216 (22.5%)	77 (27.5%)
	Post-menopausal	35,770 (54.5%)	548 (67.2%)	9,324 (62.9%)	464 (68.6%)	4,197 (77.5%)	203 (72.5%)
Age at menopause (among post-menopausal women)*, N (%)							
	<40	721 (2.0%)	12 (2.2%)	170 (1.8%)	9 (1.9%)	75 (1.8%)	5 (2.5%)
	[40,45]	1,466 (4.1%)	23 (4.2%)	368 (3.9%)	21 (4.5%)	147 (3.5%)	10 (4.9%)
	[45,50]	4,116 (11.5%)	75 (13.7%)	1,037 (11.1%)	65 (14.0%)	435 (10.4%)	32 (15.8%)
	[50,55]	7,706 (21.5%)	135 (24.6%)	1,889 (20.3%)	117 (25.2%)	842 (20.1%)	58 (28.6%)
	≥55	2,640 (7.4%)	57 (10.4%)	742 (8.0%)	49 (10.6%)	355 (8.4%)	27 (13.3%)
	Missing	19,121(53.4%)	246 (44.9%)	5,118(54.9%)	203 (43.8%)	2,343(55.8%)	71 (35.0%)
Use of hormonal replacement treatment (among post-menopausal women)*, N (%)							
	Current C	655 (1.8%)	19 (3.5%)	152 (1.6%)	14 (3.0%)	55 (1.3%)	4 (2.0%)
	Current E	806 (2.3%)	25 (4.6%)	198 (2.1 %)	18 (3.9%)	84 (2.0%)	5 (2.5%)
	Former	8,005 (22.4%)	138 (25.2%)	2,197 (23.6%)	116 (25.0%)	1,058 (25.2%)	50 (24.6%)
	Never	20,657 (57.7%)	283 (51.6%)	5,650 (60.6%)	244 (52.6%)	2,493 (59.4%)	114 (56.1%)
	Missing	5,647 (15.8%)	83 (15.1%)	1,127 (12.1%)	72 (15.5%)	507 (12.1%)	30 (14.8%)
Parity*, N (%)							
	0	7,785 (11.9%)	94 (11.5%)	1,809 (12.2%)	80 (11.8%)	609 (11.2%)	34 (12.1%)
	1	8,935 (13.6%)	127 (15.6%)	2,164 (14.6%)	108 (16.0%)	788 (14.6%)	54 (19.3%)
	2	29,380 (44.8%)	377 (46.2%)	7,007 (47.3%)	312 (46.1%)	2,562 (47.3%)	115 (41.1%)
	≥3	15,202 (23.2%)	170 (20.8%)	3,740 (25.2%)	137 (20.3%)	1,445 (26.7%)	65 (23.2%)
	Missing	4,297 (6.5%)	48 (5.9%)	106 (0.7%)	39 (5.8%)	9 (0.2%)	12 (4.3%)
Age at first live birth (among parious women)*, N (%)							
	<20	3,108 (5.8%)	42 (6.2%)	842 (6.5%)	36 (6.5%)	366 (7.6%)	15 (6.4%)
	[20,25]	14,915 (27.9%)	183 (27.2%)	3,815 (29.5%)	145 (26.0%)	1,560 (32.5%)	57 (24.3%)
	[25,30]	18,847 (35.2%)	240 (35.6%)	4,562 (35.3%)	203 (36.4%)	1,783 (37.2%)	83 (35.5%)
	≥30	16,625 (31.1%)	209 (31.0%)	3,682 (28.5%)	173 (31.1%)	1,081 (22.5%)	79 (33.8%)
	Missing	22 (0.04%)	0 (0.0%)	10 (0.1%)	0 (0.0%)	5 (0.1%)	0 (0.0%)

Use of oral contraceptive*, N (%)						
Ever	51,882 (79.1%)	632 (77.5%)	12,198 (82.3%)	526 (77.8%)	4,307 (79.6%)	221 (78.9%)
Never	8,834 (13.5%)	134 (16.4%)	2,341 (15.8%)	110 (16.3%)	1,013 (18.7%)	46 (16.4%)
Missing	4,883 (7.4%)	50 (6.1%)	287 (1.9%)	40 (5.9%)	93 (1.7%)	13 (4.6%)
Body Mass Index* (kg/m2), N (%)						
<18.5	4,298 (6.6%)	44 (5.4%)	159 (1.1%)	37 (5.5%)	53 (1.0%)	9 (3.2%)
[18.5,25)	33,912 (51.7%)	391 (47.9%)	7,878 (53.1%)	323 (47.8%)	2,700 (49.9%)	129 (46.1%)
[25,30)	19,449 (29.6%)	275 (33.7%)	4,781 (32.3%)	228 (33.7%)	1,865 (34.4%)	96 (34.3%)
≥30	7,940 (12.1%)	106 (13.0%)	2,008 (13.5%)	88 (13.0%)	795 (14.7%)	46 (16.4%)
Height* (cm), N (%)						
<152.91	625 (1.0%)	4 (0.5%)	144 (1.0%)	4 (0.6%)	55 (1.0%)	2 (0.7%)
[152.91, 159.65)	6,240 (9.5%)	67 (8.2%)	1,528 (10.3%)	53 (7.8%)	591 (10.9%)	27 (9.6%)
[159.65, 165.96)	20,366 (31.0%)	237 (29.0%)	5,014 (33.8%)	195 (28.9%)	1,870 (34.6%)	88 (31.4%)
[165.96, 172.70)	24,645 (37.6%)	333 (40.8%)	5,847 (39.4%)	280 (41.4%)	2,108 (38.9%)	111 (39.6%)
≥172.70	10,053 (15.3%)	135 (16.5%)	2,284 (15.4%)	111 (16.4%)	787 (14.5%)	43 (15.4%)
Missing	3,670 (5.6%)	40 (4.9%)	9 (0.1%)	33 (4.9%)	2 (0.0%)	9 (3.2%)
Alcohol consumption* (g/day), N (%)						
<5	11,738 (17.9%)	130 (15.9%)	2,830 (19.1%)	108 (16.0%)	1,050 (19.4%)	41 (14.6%)
[5,15)	40,501 (61.7%)	506 (62.0%)	9,670 (65.2%)	423 (62.6%)	3,484 (64.4%)	184 (65.7%)
[15,25)	5,918 (9.0%)	93 (11.4%)	1,478 (10.0%)	76 (11.2%)	599 (11.1%)	32 (11.4%)
[25,35)	1,213 (1.9%)	13 (1.6%)	290 (2.0%)	8 (1.2%)	116 (2.1%)	3 (1.1%)
[35,45)	1,195 (1.8%)	19 (2.3%)	274 (1.8%)	16 (2.4%)	103 (1.9%)	6 (2.1%)
≥45	219 (0.3%)	3 (0.4%)	57 (0.4%)	2 (0.3%)	22 (0.4%)	1 (0.4%)
Missing	4,815 (7.3%)	52 (6.4%)	227 (1.5%)	43 (6.4%)	39 (0.7%)	13 (4.6%)
BI-RADS, N (%)						
A	8,209 (12.5%)	64 (7.8%)	2,173 (14.7%)	52 (7.7%)	949 (17.5%)	28 (10.0%)
B	24,551 (37.4%)	334 (40.9%)	5,981 (40.3%)	279 (41.3%)	2,443 (45.1%)	113 (40.4%)
C	25,982 (39.6%)	329 (40.3%)	5,424 (36.6%)	272 (40.2%)	1,704 (31.5%)	116 (41.4%)
D	6,857 (10.5%)	89 (10.9%)	1,248 (8.4%)	73 (10.8%)	317 (5.9%)	23 (8.2%)
Number of 1 st degree relatives with BC, N (%)						

	0	58,102 (88.6%)	660 (80.9%)	13,050 (88.0%)	547 (80.9%)	4,783 (88.4%)	225 (80.4%)
	1	7,183 (10.9%)	145 (17.8%)	1,699 (11.5%)	124 (18.3%)	602 (11.1%)	53 (18.9%)
	≥2	314 (0.5%)	11 (1.4%)	77 (0.5%)	5 (0.7%)	28 (0.5%)	2 (0.7%)
Standardized PRS, mean (sd)							
		-	-	-0.0089 (1.03)	0.44 (1.01)	-0.068 (1.03)	0.51 (1.01)
BRCA1, N (%)							
	Negative	-	-	-	-	5,410 (99.9%)	279 (99.6%)
	Positive	-	-	-	-	3 (0.1%)	1 (0.4%)
	Untested	-	-	-	-	0 (0.0%)	0 (0.0%)
BRCA2, N (%)							
	Negative	-	-	-	-	5,403 (99.8%)	278 (99.3%)
	Positive	-	-	-	-	10 (0.2%)	2 (0.7%)
	Untested	-	-	-	-	0 (0%)	0 (0%)
PALB2, N (%)							
	Negative	-	-	-	-	5,408 (99.9%)	278 (99.3%)
	Positive	-	-	-	-	5 (0.1%)	2 (0.7%)
	Untested	-	-	-	-	0 (0.0%)	0 (0.0%)
ATM, N (%)							
	Negative	-	-	-	-	5,402 (99.8%)	278 (99.3%)
	Positive	-	-	-	-	11 (0.2%)	2 (0.7%)
	Untested	-	-	-	-	0 (0.0%)	0 (0.0%)
CHEK2, N (%)							
	Negative	-	-	-	-	5,356 (99.0%)	275 (98.2%)
	Positive	-	-	-	-	57 (1.0%)	5 (1.8%)
	Untested	-	-	-	-	0 (0%)	0 (0%)
RAD51C, N (%)							
	Negative	-	-	-	-	5,408 (99.9%)	280 (100.0%)
	Positive	-	-	-	-	5 (0.1%)	0 (0.0%)
	Untested	-	-	-	-	0 (0.0%)	0 (0.0%)
RAD51D, N (%)							

BARD1, N (%)	Negative	-	-	-	-	5,413 (100.0%)	279 (99.6%)
	Positive	-	-	-	-	0(0.0%)	1 (0.4%)
	Untested	-	-	-	-	0 (0.0%)	0 (0.0%)
	Negative	-	-	-	-	5,356 (99.9%)	280 (100.0%)
	Positive	-	-	-	-	6 (0.1%)	0 (0.0%)
	Untested	-	-	-	-	0 (0.0%)	0 (0.0%)

* These are questionnaire based risk factors (QRFss), considered in BOADICEA. Risk factor categories as defined in the BOADICEA model.

- Not applicable

Table s3 Tumour characteristics for incident breast cancer patients.

	Full cohort (N=816)	Sub cohort with PRS (N=676)	Sub cohort with PRS and PV (N=280)
Stage, N (%)			
0	4 (0.5%)	4 (0.6%)	3 (1.1%)
1	444 (54.4%)	357 (52.8%)	146 (52.1%)
2	216 (26.5%)	182 (26.9%)	69 (24.6%)
3	74 (9.1%)	65 (9.6%)	31 (11.1%)
4	6 (0.7%)	5 (0.7%)	1 (0.4%)
missing	72 (8.8%)	63 (9.3%)	30 (10.7%)
Morphology, N (%)			
ductal	622 (76.2%)	516 (76.3%)	218 (77.9%)
lobular	97 (11.9%)	83 (12.3%)	39 (13.9%)
Medullary	2 (0.2%)	1 (0.1%)	0 (0.0%)
mixed (ductal and lobular)	13 (1.6%)	11 (1.6%)	3 (1.1%)
missing	82 (10.0%)	65 (9.6%)	20 (7.1%)
ER status, N (%)			
negative	97 (11.9%)	80 (11.8%)	31 (11.1%)
positive	671 (82.2%)	558 (82.5%)	232 (82.9%)
missing	48 (5.9%)	38 (5.6%)	17 (6.1%)

Table s4 Reclassification table comparing the model considering family history (FH), questionnaire-based risk factors (QRFs), mammographic density (MD) in BI-RADS and polygenic risk score (PRS) and the model considering FH using the subcohort samples with PRS. The table shows the weighted number of women in each category of risk (under each model). The weighted number of incident breast cancer patients is shown in the brackets.

Model: FH	Model: FH+QRFs+MD+PRS		
	Five-year risk < 3%	Five-year risk ≥ 3%	Total
All women			
Five-year risk < 3%	64049 (733)	1894 (68)	65943 (801)
Five-year risk ≥ 3%	332 (7)	189 (7)	521 (14)
Total	64381 (740)	2083 (75)	66464 (815)
Premenopausal women			
Five-year risk < 3%	30027 (239)	722 (23)	30749 (262)
Five-year risk ≥ 3%	9 (0)	30 (1)	39 (1)
Total	30036 (239)	752 (24)	30788 (263)
Postmenopausal women			
Five-year risk < 3%	34022 (494)	1172 (45)	35194 (539)
Five-year risk ≥ 3%	323 (7)	159 (6)	482 (13)
Total	34345 (501)	1331 (51)	35676 (552)

Table s5 Sensitivity and specificity at different risk thresholds and proportions of ER-negative and ER-positive cancers detected in each risk category.

Risk threshold/ category	Model: FH+QRFs+MD+PRS [†]				Model: FH+QRFs+MD+PRS+PV [‡]			
	BCs=676		BCs with ER status=644		BCs=280		BCs with ER status=267	
	Sensitivity	Specificity	ER-negative cancers (%*)	ER-positive cancers (**)	Sensitivity	Specificity	ER-negative cancers (%*)	ER-positive cancers (**)
<1.10%	-	-	34 (42.5%)	166 (29.4%)	-	-	15 (48.4%)	69 (29.2%)
[1.10 [#] -1.67%)	0.69 (0.66, 0.72)	0.59 (0.59, 0.59)	22 (27.5%)	181 (32.1%)	0.66 (0.62, 0.69)	0.63 (0.63, 0.63)	5 (16.1%)	72 (30.5%)
[1.67 [#] -3.00%)	0.38 (0.34, 0.41)	0.81 (0.81, 0.82)	22 (27.5%)	161 (28.5%)	0.34 (0.31, 0.37)	0.83 (0.83, 0.84)	8 (25.8%)	70 (29.7%)
≥3% [#]	0.09 (0.01, 0.11)	0.97 (0.97, 0.97)	2 (2.5%)	56 (9.9%)	0.09 (0.07, 0.11)	0.97 (0.96, 0.97)	3 (9.7%)	25 (10.6%)

[#]Risk thresholds (≥) for calculating sensitivity/specificity

*among ER-negative; **among ER-positive

[†]Using the weighted sub-cohort with PRS (N=15,502); [‡]Using the weighted sub-cohort with PRS and PV (N=5,693). Therefore, the results were not directly comparable due to the different sample sizes.

FH: family history; QRFs: questionnaire-based risk factors; MD: mammographic density in BI-RADS; PRS: polygenic risk score; PV: pathogenic variants in *BRCA1*, *BRCA2*, *PALB2*, *CHECK2*, *ATM*, *RAD51C*, *RAD51D*, and *BARD1*.

Table s6 Sensitivity and specificity at different percentiles of the predicted risk distribution, under the model considering family history (FH), questionnaire-based risk factors (QRFs), mammographic density in BI-RADS (MD) and polygenic risk score (PRS).

Risk distribution threshold	Equivalen 5-year risk threshold	sensitivity	specificity
Top 2.5%	3.26%	0.07 (0.05, 0.09)	0.98 (0.98, 0.98)
Top 5%	2.72%	0.13 (0.11, 0.15)	0.96 (0.95, 0.96)
Top 10%	2.20%	0.23 (0.20, 0.26)	0.91 (0.91, 0.91)
Top 20%	1.69%	0.37 (0.34, 0.41)	0.82 (0.82, 0.82)
Top 30%	1.39%	0.53 (0.49, 0.56)	0.73 (0.72, 0.73)
Top 40%	1.18%	0.65 (0.61, 0.68)	0.63 (0.63, 0.63)
Top 50%	1.00%	0.75 (0.72, 0.78)	0.53 (0.53, 0.54)

Table s7 Reclassification table comparing the model that considers family history (FH), questionnaire-based risk factors (QRFs), mammographic density (MD) in BI-RADS, polygenic risk score (PRS) and pathogenic variants (PV) in the five major breast cancer (BC) susceptibility genes and the model considering FH, QRFs, MD and PRS, based on the subcohort samples with PRS and PV status. The table shows the weighted number of women in each category of risk. The weighted number of incident BC patients is shown in the brackets.

Model: FH+QRFs+MD+PRS	Model: FH+QRFs+MD+PRS+PV		
	Five-year risk < 3%	Five-year risk ≥ 3%	Total
Five-year risk < 3%	63412 (736)	647 (23)	64059 (759)
Five-year risk ≥ 3%	228 (7)	1702 (49)	1930 (59)
Total	63640 (743)	2349 (67)	65989 (816)

Table s8 Risk reclassification of the 110 pathogenic variant (PV) carriers using five-year predicted breast cancer risk under different models. FH: family history; QRFs: questionnaire-based risk factors; MD: mammographic density in BI-RADS; PRS: polygenic risk score; PV: pathogenic variants in *BRCA1*, *BRCA2*, *PALB2*, *CHECK2*, *ATM*, *RAD51C*, *RAD51D*, and *BARD1*.

Model	Risk <1.67%	Risk: 1.67-3%	Risk: ≥3%	E/O
FH Only	45%	53%	2%	0.35
+ PRS+ MD+ QRFs	68%	27%	5%	0.31
+ PRS+ MD+ QRFs+ PV	25%	23%	52%	0.99
PV only	3%	36%	61%	0.99
+ PRS+MD+QRFs+FH	25%	23%	52%	0.99

Table s9 Calibration and discrimination of five-year predicted breast cancer risks in women 40 years old or older using different risk factor combinations.

Model	AUC (95%CI)	Harrell's C (95%CI)	E/O (95%CI)	Calibration slope (95%CI)
Entire cohort with information on FH, QRF and MD (N=66,152; N.BCs=816)				
FH	0.62 (0.60, 0.64)	0.62 (0.61, 0.65)	1.12 (1.05, 1.20)	1.03 (1.01, 1.05)
QRF	0.62 (0.60, 0.64)	0.62 (0.60, 0.64)	0.95 (0.89, 1.02)	0.99 (0.97, 1.01)
MD	0.62 (0.60, 0.64)	0.62 (0.60, 0.64)	0.94 (0.88, 1.01)	0.99 (0.97, 1.01)
FH+QRF	0.63 (0.61, 0.65)	0.63 (0.61, 0.65)	1.02 (0.95, 1.09)	1.00 (0.99, 1.02)
FH+QRF+MD	0.64 (0.62, 0.66)	0.63 (0.62, 0.66)	0.92 (0.86, 0.99)	0.98 (0.96, 1.00)
Sub-cohort with information on FH, QRF, MD and PRS (N=15,427; N.BCs=676)				
FH	0.61 (0.59, 0.63)	0.62 (0.60, 0.64)	1.12 (1.04, 1.21)	1.03 (1.01, 1.05)
QRF	0.62 (0.60, 0.65)	0.63 (0.61, 0.65)	0.94 (0.87, 1.01)	0.99 (0.97, 1.00)
MD	0.62 (0.60, 0.64)	0.64 (0.61, 0.66)	0.93 (0.86, 1.01)	0.98 (0.97, 1.00)
PRS	0.66 (0.64, 0.68)	0.67 (0.65, 0.69)	1.06 (0.98, 1.15)	1.01 (1.00, 1.03)
FH+QRF+MD	0.65 (0.63, 0.67)	0.65 (0.63, 0.67)	0.90 (0.84, 0.97)	0.98 (0.96, 1.00)
FH+QRF+PRS	0.68 (0.66, 0.70)	0.68 (0.66, 0.70)	1.01 (0.93, 1.09)	1.00 (0.98, 1.02)
FH+QRF+PRS+MD	0.69 (0.67, 0.71)	0.68 (0.67, 0.70)	0.91 (0.85, 0.98)	0.98 (0.96, 1.00)
Sub-cohort with information on FH, QRF, MD, PRS and PV status (N=5,653; N.BCs=280)				
FH+QRF+PRS+MD	0.68 (0.64, 0.72)	0.7 (0.66, 0.72)	0.88 (0.74, 1.04)	0.97 (0.95, 0.99)
FH+QRF+PRS+MD+PV	0.69 (0.66, 0.73)	0.7 (0.67, 0.73)	0.88 (0.75, 1.04)	0.97 (0.95, 0.99)

BC: breast cancer; E: expected number of breast cancers in the five-year period; O: observed number breast cancers; FH: family history; QRFs: questionnaire-based risk factors; MD: mammographic density in BI-RADS; PRS: polygenic risk score; PV: pathogenic variants in *BRCA1*, *BRCA2*, *PALB2*, *CHECK2*, *ATM*, *RAD51C*, *RAD51D*, and *BARD1*.

Table s10 Calibration and discrimination of five-year predicted breast cancer risks by menopausal status. Analyses restricted to women 40 years old or older.

Menopausal status	N. unaffected	N. BCs	Model	AUC (95%CI)	Harrell's C (95%CI)	E/O (95%CI)	Calibration (95%CI)	slope
Sub-cohort with information on FH, QRFs, MD, and PRS (N=15,427; N.BCs=676)								
Pre-menopausal	5,427	212	FH+QRFs+PRS	0.67 (0.63, 0.71)	0.66 (0.62, 0.7)	1.09 (0.95, 1.25)	1.02 (0.99, 1.05)	
			FH+QRFs+PRS+MD	0.68 (0.65, 0.72)	0.68 (0.65, 0.71)	1.13 (0.99, 1.3)	1.03 (0.99, 1.06)	
Post-menopausal	9,324	464	FH+QRFs+PRS	0.66 (0.63, 0.68)	0.65 (0.63, 0.68)	0.97 (0.89, 1.06)	0.99 (0.97, 1.02)	
			FH+QRFs+PRS+MD	0.67 (0.64, 0.69)	0.67 (0.64, 0.68)	0.81 (0.74, 0.89)	0.95 (0.93, 0.97)	
Sub-cohort with information on FH, QRFs, MD, PRS, and PV status (N=5,653; N.BCs=280)								
Pre-menopausal	1,176	77	FH+QRFs+PRS+PV	0.69 (0.6, 0.78)	0.67 (0.59, 0.73)	1.1 (0.77, 1.58)	1.02 (0.99, 1.05)	
			FH+QRFs+PRS+PV+MD	0.68 (0.6, 0.75)	0.67 (0.59, 0.74)	1.11 (0.78, 1.6)	1.02 (0.98, 1.05)	
Post-menopausal	4,197	203	FH+QRFs+PRS+PV	0.68 (0.63, 0.73)	0.69 (0.65, 0.72)	0.95 (0.79, 1.13)	0.99 (0.96, 1.01)	
			FH+QRFs+PRS+PV+MD	0.68 (0.64, 0.73)	0.69 (0.66, 0.73)	0.78 (0.66, 0.94)	0.94 (0.92, 0.96)	

BC: Breast Cancer; FH: family history; QRFs: questionnaire-based risk factors; MD: mammographic density in BI-RADS; PRS: polygenic risk score; PV: pathogenic variants in *BRCA1*, *BRCA2*, *PALB2*, *CHECK2*, *ATM*, *RAD51C*, *RAD51D*, and *BARD1*.

Figure s1 Predicted five-year risk distribution for the incident breast cancer patients (affected) and healthy women (unaffected) under different models using the entire cohort (N=66,415). The box represents the interquartile range (IQR) and the central line within the box represents the mean. The x-axis is on the log-scale.

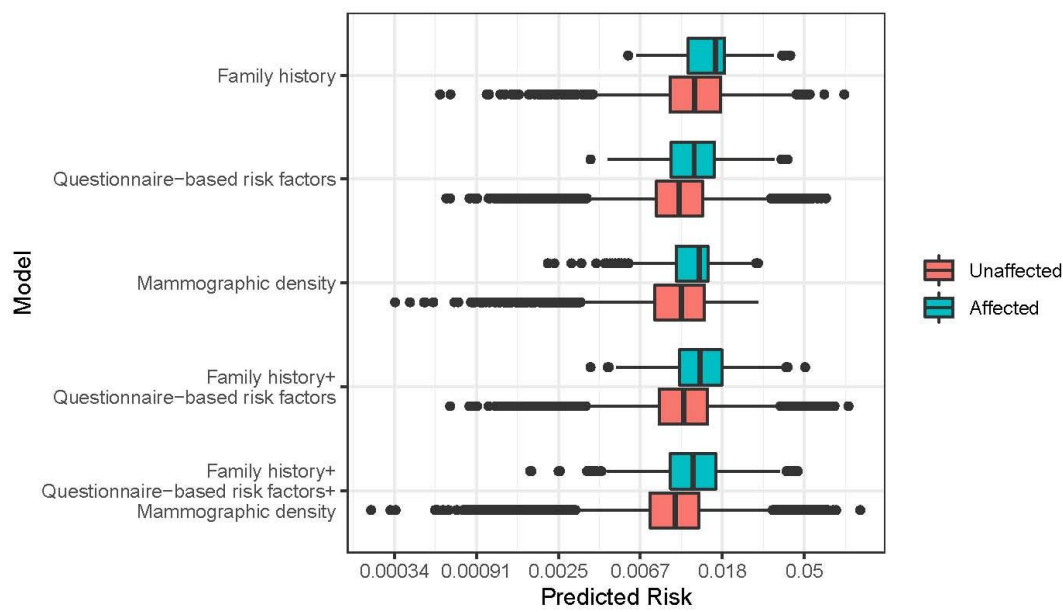


Figure s2 Observed and predicted five-year breast cancer risks in deciles of predicted risks using the entire cohort under different models Women were grouped into deciles of predicted risks. Each dot represents the mean observed and predicted risk in the decile and the vertical segments represent 95% confidence intervals. The dashed line is the diagonal line with slope equal to 1 (corresponding to E/O ratio of 1 for each decile). When the confidence interval crosses the diagonal, the decile-predicted risk is not significantly different from the observed risk. When a dot and the associated confidence interval fall above the diagonal, there is a suggestion for underprediction of risk; when a dot and associated confidence interval fall below the diagonal there is a suggestion for overprediction of risk.

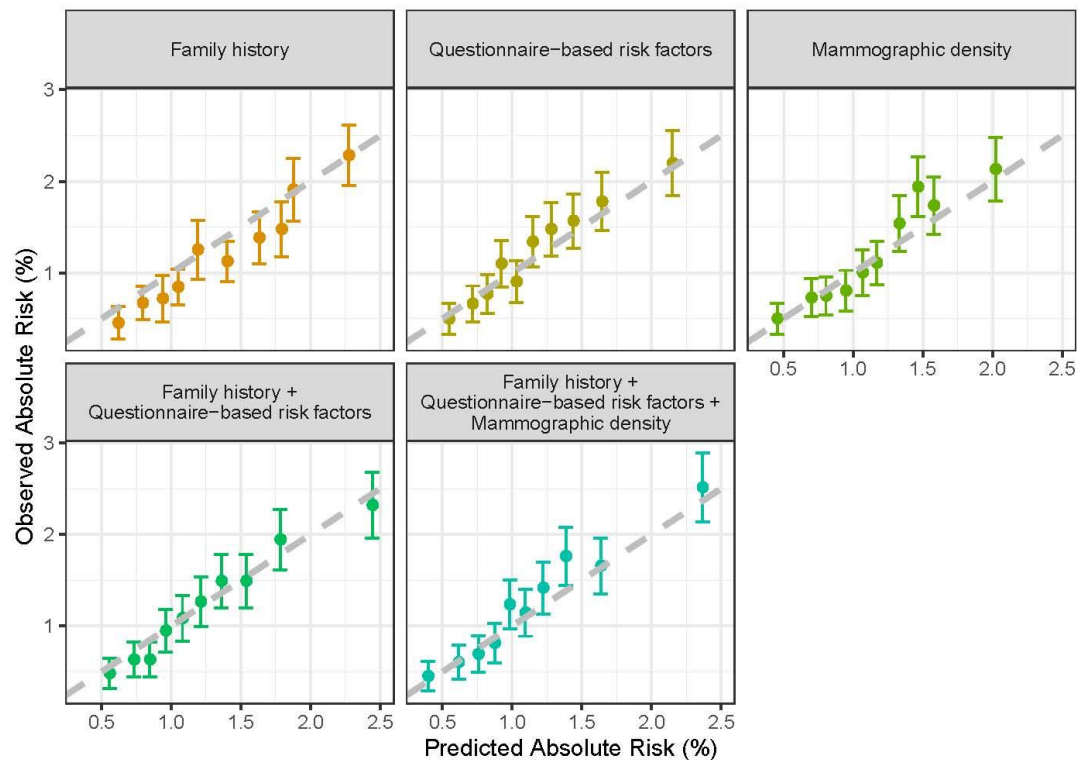


Figure s3 Predicted five-year risk distribution for the incident breast cancer patients (affected) and healthy women (unaffected) under different models using the subcohort samples with polygenic risk score (PRS) information(N=15,502). The box represents the interquartile range (IQR) and the central line within the box represents the mean. . The x-axis is on the log-scale.

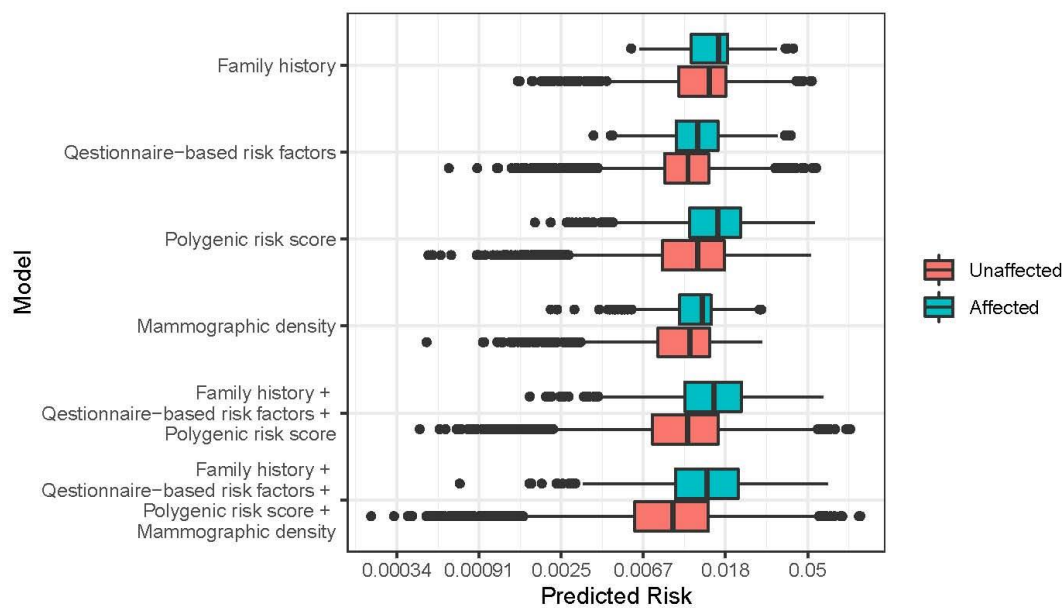


Figure s4 Predicted five-year risk distribution for the incident breast cancer patients (affected) and healthy women (unaffected) under different models using the subcohort samples with polygenic risk score (PRS) and pathogenic variants (PV) information(N=5,693). The box represents the interquartile range (IQR) and the central line within the box represents the mean. The x-axis is on the log-scale.

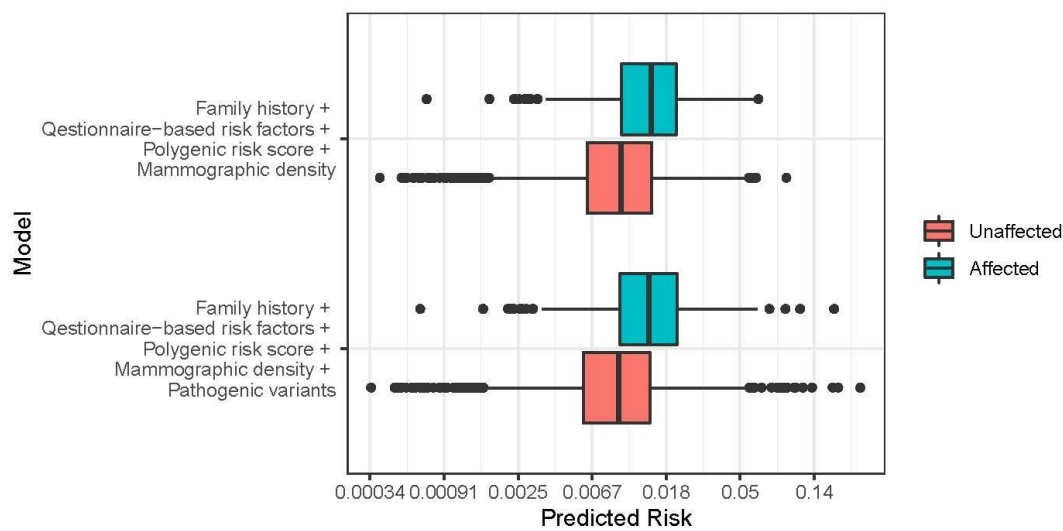


Figure s5 Predicted five-year risk distributions for carriers and non-carriers of pathogenic variants (PV) in the eight major breast cancer (BC) susceptibility genes: *BRCA1*, *BRCA2*, *PALB2*, *CHECK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1* under the models considering family history, questionnaire-based risk factor, polygenic risk score, mammographic density in BI-RADS and pathogenic variant status using the subcohort samples with PRS and PV information (N=5,693). The box represents the interquartile range (IQR) and the central line within the box represents the mean. . The x-axis is on the log-scale.

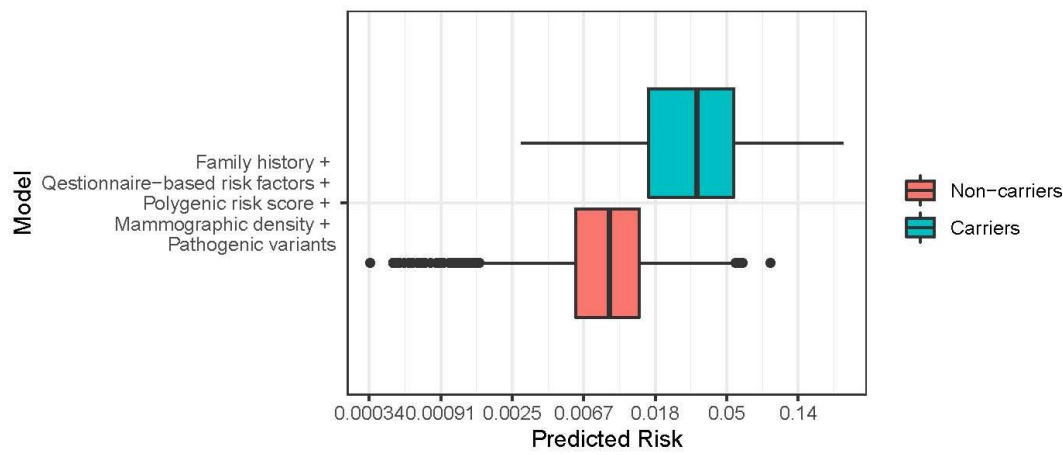


Figure s6 The distribution of BIRADS categories across deciles of predicted risks in postmenopausal women in the subcohort samples with polygenic risk score (PRS) information (N=9,788). Each bar represents the proportion of women with BIRADS A, B, C, or D among total number of unaffected or affected women. Affected: incident breast cancer patients during the five-year risk prediction horizon; Unaffected: women without a breast cancer diagnosis during the five-year risk prediction horizon.

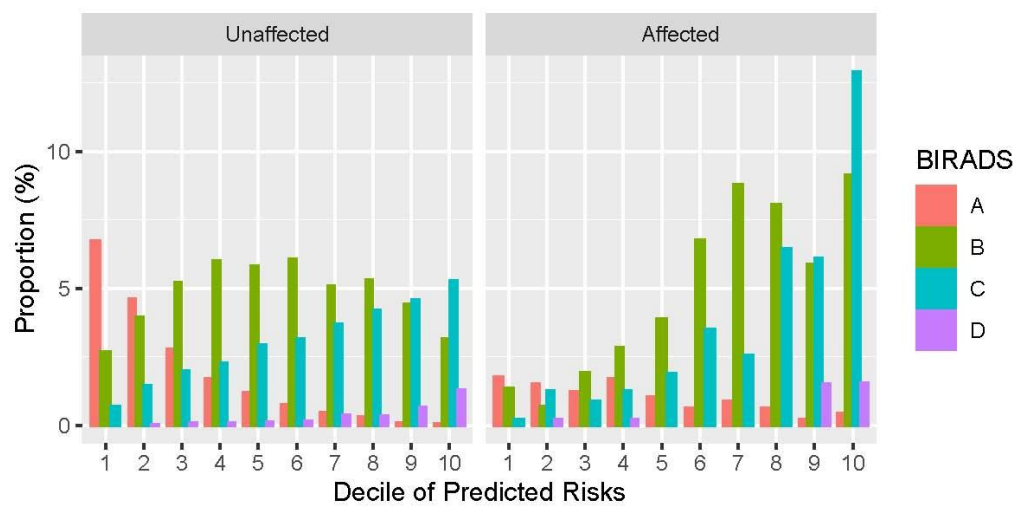


Figure s7. Observed and expected five-year breast cancer risks in deciles of predicted risks in women 40 years old or older. The sub-cohort of samples with PRS information (N=15,427) was used here under the model considering considering family history, questionnaire-based risk factors, polygenic risk score and mammographic density in BI-RADS. Women were grouped into deciles of predicted risks.

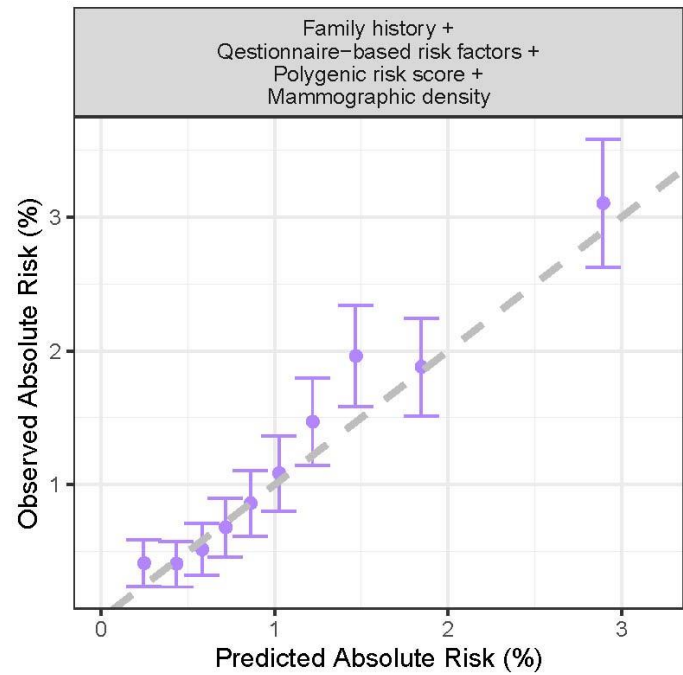


Figure s8. Observed and expected five-year breast cancer risks in deciles of predicted risks in women 40 years old or older. The sub-cohort of samples with PRS and PV status information ($N=5,653$) was used here under the model considering all risk factors including family history, questionnaire-based risk factors, polygenic risk score, mammographic density in BI-RADS and pathogenic variant status in the eight major breast cancer (BC) susceptibility genes: *BRCA1*, *BRCA2*, *PALB2*, *CHECK2*, *ATM*, *RAD51C*, *RAD51D* and *BARD1*. Women were grouped into deciles of predicted risks.

