




ARTICLE

Performance of Breast Cancer Risk-Assessment Models in a Large Mammography Cohort

Anne Marie McCarthy*, Zoe Guan*, Michaela Welch, Molly E. Griffin, Dorothy A. Sippo , Zhengyi Deng, Suzanne B. Coopey, Ahmet Acar, Alan Semine, Giovanni Parmigiani, Danielle Braun , Kevin S. Hughes 

See the Notes section for the full list of authors' affiliations.

*Authors contributed equally to this work.

Correspondence to: Anne Marie McCarthy, ScM, PhD, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Room 833, 423 Guardian Drive, Philadelphia, PA 19004 (e-mail: annemcc@pennmedicine.upenn.edu).

Abstract

Background: Several breast cancer risk-assessment models exist. Few studies have evaluated predictive accuracy of multiple models in large screening populations.

Methods: We evaluated the performance of the BRCAPRO, Gail, Claus, Breast Cancer Surveillance Consortium (BCSC), and Tyrer-Cuzick models in predicting risk of breast cancer over 6 years among 35 921 women aged 40–84 years who underwent mammography screening at Newton-Wellesley Hospital from 2007 to 2009. We assessed model discrimination using the area under the receiver operating characteristic curve (AUC) and assessed calibration by comparing the ratio of observed-to-expected (O/E) cases. We calculated the square root of the Brier score and positive and negative predictive values of each model.

Results: Our results confirmed the good calibration and comparable moderate discrimination of the BRCAPRO, Gail, Tyrer-Cuzick, and BCSC models. The Gail model had slightly better O/E ratio and AUC (O/E = 0.98, 95% confidence interval [CI] = 0.91 to 1.06, AUC = 0.64, 95% CI = 0.61 to 0.65) compared with BRCAPRO (O/E = 0.94, 95% CI = 0.88 to 1.02, AUC = 0.61, 95% CI = 0.59 to 0.63) and Tyrer-Cuzick (version 8, O/E = 0.84, 95% CI = 0.79 to 0.91, AUC = 0.62, 95% CI = 0.60 to 0.64) in the full study population, and the BCSC model had the highest AUC among women with available breast density information (O/E = 0.97, 95% CI = 0.89 to 1.05, AUC = 0.64, 95% CI = 0.62 to 0.66). All models had poorer predictive accuracy for human epidermal growth factor receptor 2 positive and triple-negative breast cancers than hormone receptor positive human epidermal growth factor receptor 2 negative breast cancers.

Conclusions: In a large cohort of patients undergoing mammography screening, existing risk prediction models had similar, moderate predictive accuracy and good calibration overall. Models that incorporate additional genetic and nongenetic risk factors and estimate risk of tumor subtypes may further improve breast cancer risk prediction.

Approximately 40 000 US women die from breast cancer annually (1). Given the disease burden, identifying high-risk women before breast cancer develops remains a pressing goal, so they can consider more frequent screening with mammography and breast magnetic resonance imaging, genetic testing, and chemoprevention. Many breast cancer risk-assessment models have been developed (2). Despite the abundance of risk models, they have not been widely implemented to guide screening

decisions in routine clinical settings. This is partly due to lack of clarity on which risk model to use, limited accuracy of risk models, and the time needed to perform risk assessment and interpret results.

Few studies have evaluated multiple breast cancer risk models simultaneously to compare their performance. Additionally, breast cancer subtypes defined by estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor 2

Received: March 28, 2019; Revised: July 23, 2019; Accepted: September 4, 2019

© The Author(s) 2019. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

(HER2) have unique risk profiles (3), but few studies have evaluated model performance for tumor subtypes. We evaluated five models that have been well validated and used most commonly in clinical practice in a large population of women undergoing mammography, including the Gail model (4–7), the BRCAPRO model (8), the Breast Cancer Surveillance Consortium (BCSC) model (9), the Claus model (10–12), and the Tyrer-Cuzick (TC) model (13). These models differ in terms of populations in which they were developed, risk factors included, and treatment of family history (Table 1). The goal of the study was to determine which risk models are most appropriate for use at the time of screening mammography to guide personalized screening decisions.

Methods

Study Population and Risk Assessment

We retrospectively assembled a cohort of women presenting to Newton-Wellesley Hospital (Newton, MA) for mammography between February 1, 2007, and December 31, 2009. The Partners Healthcare Institutional Review Board approved the study and waived the need for informed consent. Patient-reported risk factors were collected prior to mammography using an electronic questionnaire through the Hughes RiskApps system (CRA Health: <http://www.crahealth.com/>). If a woman had multiple risk assessments, only her first was included. A total of 42 919 women underwent mammography and risk assessment during the time frame. Women aged younger than 40 years ($n=4352$) and older than 84 years ($n=744$) were excluded. Patients diagnosed with ductal carcinoma in situ (DCIS) or invasive breast cancer before risk assessment ($n=1705$) or within 6 months of a positive or unknown mammogram result ($n=165$) were excluded ($n=1893$) because most models predict risk for women who are breast cancer-free at the time of risk assessment. A sensitivity analysis excluding all patients diagnosed with breast cancer within 6 months of screening is included in the Supplementary Materials (Supplementary Table 1). Additionally, women with lobular carcinoma in situ (LCIS) before risk assessment ($n=15$), women who tested positive for a BRCA1/2 mutation before risk assessment ($n=14$), and women with inconsistent dates of screening and death ($n=3$) were excluded, yielding a study population of 35 921. Breast Imaging Reporting and Data System (BI-RADS) assessment and mammographic breast density were abstracted from radiology report text using JMP software (Version 14, SAS Institute, Cary, NC). We compared the abstracted Breast Imaging Reporting and Data System score and density to the results of a manual review of 300 radiology reports and found excellent agreement ($\kappa=0.99$; $\kappa=0.97$), validating use of this tool.

Outcome Assessment

Breast cancer diagnoses through December 31, 2015, were determined from three sources: the Massachusetts Cancer Registry (MCR), the cancer registries of Newton-Wellesley Hospital and five nearby affiliated hospitals, and from self-report verified by medical record abstraction. Nearly 82% of cancer cases were obtained from the Massachusetts Cancer Registry ($n=629$, 81.6%), an additional 6.0% from hospital registries ($n=46$), and 12.3% from self-report/medical record abstraction ($n=95$). Most of the self-reported cases were DCIS (92 out of 95).

Statistical Analysis

Patients were followed from the date of mammography until cancer diagnosis, death, or administrative censoring on December 31, 2015, allowing at least 6 years of follow-up. We estimated 6-year risk of breast cancer using BRCAPRO, Gail, TC versions 7 and 8 (v7 and v8), BCSC, and Claus models, and compared with observed breast cancer outcomes. Death was treated as a competing risk. We used the BayesMendel R package (v2.1–4) (14); BCRA R package (v2.1) (15); IBIS command line program (v8.0b, UK rates with competing mortality option) (16); and BCSC SAS program (v2.0) (17) to run BRCAPRO, Gail, TC, and BCSC, respectively. For Claus, we used R to calculate risk estimates based on the Claus risk tables (11,12). Because the Claus model provides predictions for 10-year intervals from ages 29 to 79 years, to obtain 6-year risk estimates, we multiplied the 10-year risk by 3/5 (18). Similarly, BCSC provides 5-year breast cancer risk, and 6-year risk estimates were obtained by multiplying the risk estimate by 6/5.

Missing data were coded according to the specifications of the software for each model. Information on relatives unaffected by breast/ovarian cancer (used in BRCAPRO and TC) was not available unless they had been diagnosed with another cancer. There was also limited genetic testing information and no information on bilateral breast cancer (used in BRCAPRO and TC). Polygenic risk scores (used in TC v8) were not available. For the Claus model, missing breast/ovarian cancer diagnosis ages among relatives (5.0%) were set to age 50 years. To account for combinations of affected female first- or second-degree relatives that do not have a corresponding Claus risk table, we used the mother-maternal aunt table for any combination of a first-degree relative and a maternal second-degree relative and we used the mother-paternal aunt table for any combination of a first-degree relative and a paternal second-degree relative. For the BCSC model, women with unknown breast density ($n=2249$) were excluded.

We validated the performance of BRCAPRO, Gail, TC, and BCSC with respect to risk of invasive breast cancer, and Claus with respect to risk of both DCIS and invasive breast cancer. We compared the performance of BRCAPRO, Gail, and TC models on the entire study cohort. We additionally compared BRCAPRO, Gail, and TC with the BCSC model on the subset of patients for whom BCSC was applicable (<75 years, known breast density, no mastectomy, and no breast implants) and with the Claus model on the subset of patients for whom Claus was applicable (age <79 years, ≥ 1 female first- or second-degree relative with breast cancer or ≥ 1 first-degree relative with ovarian cancer). We performed stratified analyses by family history, Ashkenazi Jewish ancestry, age, and invasive tumor subtypes defined by ER, PR, and HER2. We also calculated the Pearson correlation between risk predictions from each pair of models.

Discrimination was assessed using the area under the receiver operating characteristic curve (AUC). An AUC of 1 corresponds to perfect discrimination, whereas an AUC of 0.5 indicates that the model performs no better than chance. Calibration was assessed by comparing the ratio of observed-to-expected cases. An observed-to-expected ratio (O/E) of 1 corresponds to perfect calibration. Additionally, we plotted the expected risk vs observed proportion of invasive breast cancer cases in each decile of risk. We also calculated the square root of the Brier score, which is the average squared difference between the observed outcome and the predicted risk. A lower Brier score indicates better model accuracy. The recognized clinical threshold for elevated 5-year risk of breast cancer for chemoprevention is 1.67%. For our study, we considered a

Table 1. Comparison of breast cancer risk models*

Model characteristic	BRCAPRO	Gail/BCRAT	Tyrer-Cuzick/IBIS	BCSC	Claus
Model development	Genetic model parameters (BRCA mutation prevalence and penetrance) estimated from meta-analysis of published studies; for non-carriers, calibrated to US incidence rates	Relative risks estimated from case-control study of US mammography screening population; calibrated to US incidence rates	Genetic model parameters and relative risks for non-genetic factors based on published studies; calibrated to UK incidence rates	Hazard ratios estimated from population-based mammography screening cohort in the United States; calibrated to US incidence rates	Genetic model parameters estimated via segregation analysis of population-based case-control study of young breast cancer patients in the United States; calibrated to US incidence rates
Outcomes predicted	Invasive breast cancer	Invasive breast cancer	Invasive breast cancer	Invasive breast cancer	DCIS or invasive breast cancer
Age, y	Any	35–90	19–85	35–74	20–79
Hormonal factors					
Age at menarche, y	No	Yes	Yes	No	No
Age at first live birth, y	No	Yes	Yes	No	No
Age at menopause, y	No	No	Yes	No	No
Parity	No	No	Yes	No	No
Hormone replacement therapy	No	No	Yes	No	No
Demographics					
Race/ethnicity	Yes	Yes	Yes	Yes	No
Ashkenazi Jewish ancestry	Yes	No	Yes	No	No
Breast pathology					
Prior/number of breast biopsies	No	Yes	Yes	Yes	No
Atypical hyperplasia	No	Yes	Yes	Yes	No
LCIS	No	Yes	Yes	Yes	No
Family history					
Relatives (degree)	Any	First	First, second, and third	First	First and second
Age at onset	Yes	No	Yes	No	Yes
Bilateral breast cancer	Yes	No	Yes	No	No
Ovarian cancer (personal or familial)	Yes	No	Yes	No	Yes
Breast density	No	No	Yes	Yes	No
Exclusions	None	Prior breast cancer, DCIS, LCIS, chest irradiation, BRCA1/2 mutation	None	Prior breast cancer, DCIS, mastectomy, breast implants, no breast density measurement	No first- or second-degree relatives with breast cancer or first-degree relatives with ovarian cancer
Prediction period	Any number of years up to age 110 y	Any number of years up to age 90 y	1–20 y or lifetime risk to age 85 y	5 y or 10 y	10-y intervals

*BCSC = Breast Cancer Surveillance Consortium; BRCA = breast cancer type; DCIS = ductal carcinoma in situ; LCIS = lobular carcinoma in situ.

Table 2. Cohort characteristics

Variable	No. (%)
N	35 921
Age, mean (SD), y	53.9 (10.6)
Age category, y	
40–49	15 230 (42.4)
50–59	11 203 (31.2)
60–69	5842 (16.3)
70+	3646 (10.2)
Race	
White	29 641 (82.5)
Black	467 (1.3)
Hispanic	482 (1.3)
Asian	1088 (3.0)
Native American	22 (0.1)
Other/Unknown	4221 (11.8)
Ashkenazi Jewish	
Yes	5165 (14.4)
No/Unknown	30 756 (85.6)
Relatives with breast cancer, categorized	
0	24 124 (67.2)
1	8823 (24.6)
2+	2974 (8.3)
Relatives with ovarian cancer, categorized	
0	33 908 (94.4)
1+	2013 (5.6)
BI-RADS density*	
1	744 (2.1)
2	9234 (25.7)
3	18 988 (52.9)
4	4706 (13.1)
Missing	2249 (6.3)
Years of follow-up, mean (interquartile range)	6.7 (6.3 to 7.1)

*Breast Imaging Reporting and Data System (BI-RADS) breast density category 1 = almost entirely fat; 2 = scattered fibroglandular density; 3 = heterogeneously dense; and 4 = extremely dense.

corresponding 6-year risk threshold of 2.0% (ie, $1.67\% \times 6/5$). Positive and negative predictive values of the models were calculated relative to this risk threshold. We calculated 95% bootstrap percentile confidence interval (CIs) for the performance metrics. All analyses were performed using R statistical software (www.R-project.org).

Results

The mean age of the study population was 53.9, and 42.4% of women were aged in their 40s (Table 2). Most patients were white (82.5%), and 14.4% had Ashkenazi Jewish ancestry. Of the women, 32.9% reported a family history of breast cancer, and 5.6% reported a family history of ovarian cancer. The mean follow-up time was 6.7 years (interquartile range = 6.3–7.1 years). There were 770 (2.1%) breast cancers diagnosed, of which 123 were DCIS and 647 were invasive, and 909 deaths (2.5%) within 6 years of follow-up.

The performance of BRCAPRO, Gail, TC v7 and TC v8 was compared in the entire study population (Table 3; receiver operating characteristic curves are in Supplementary Table 1). The Gail model had slightly better calibration (O/E = 0.98, 95% CI = 0.91 to 1.06) than BRCAPRO (O/E = 0.94, 95% CI = 0.88 to 1.02) and TC (v7 O/E = 0.90, 95% CI = 0.84 to 0.96; v8 O/E = 0.84, 95% CI = 0.79 to 0.91) did. TC v7 and v8 slightly overpredicted the number

of cancers. When examining the calibration plots of predicted and observed risk by deciles (Figure 1), the models appeared well calibrated except in the highest decile for BRCAPRO, Gail, and TC v8, and the highest two deciles for TC v7. Overprediction of risk was most severe for TC v8, which, in comparison to TC v7, uses mammographic density.

All models showed moderate discriminatory accuracy. The Gail model had the highest discriminatory accuracy (AUC = 0.64, 95% CI = 0.61 to 0.65), followed by TC v8 (AUC = 0.62, 95% CI = 0.60 to 0.64) and BRCAPRO (AUC = 0.61, 95% CI = 0.59 to 0.63) and TC v7 (AUC = 0.61, 95% CI = 0.59 to 0.63). Comparing TC v7 to TC v8, the AUC was only slightly improved with the inclusion of breast density. Using the high-risk threshold of 2.0% for 6-year risk, the positive predictive values of the models were low: 2.9% (95% CI 2.6 to 3.2%) for Gail and 2.6% (95% CI = 2.3 to 2.9) for BRCAPRO, TC v7 (95% CI = 0.024 to 0.029), and v8 (95% CI = 0.024 to 0.028). Negative predictive values were all above 98%.

When the population was limited to women with available BCSC model risk scores (N = 30 970), again, TC v7 and v8 were more prone to overprediction than the other models were, but performance metrics were otherwise similar across models. The BCSC model had the best calibration (O/E = 0.97, 95% CI = 0.89 to 1.05) and AUC (AUC = 0.64, 95% CI = 0.62 to 0.66) (Table 3).

Next, we assessed model performance stratified by family history of breast and ovarian cancer (Table 4). BRCAPRO overpredicted the number of cancers in women with no affected first- or second-degree family members, was well calibrated for women with one affected first- or second-degree family member, and underpredicted the number of cancers in women with more than two affected first- or second-degree family members. The AUC for BRCAPRO was similar for all three family history strata, ranging from 0.59 (95% CI = 0.57 to 0.62) to 0.61 (95% CI = 0.57 to 0.65). The Gail model was well calibrated for women with no family history and women with one affected family member, but it underpredicted risk among women with two or more affected family members. The AUC for the Gail model ranged from 0.61 (95% CI = 0.57 to 0.65) to 0.63 (95% CI = 0.58 to 0.66) for the three family history strata. TC v7 and v8 were well calibrated for women with no family history, but they overpredicted the number of cancers in women with one affected family member. TC v7 was well calibrated for women with two or more affected family members, whereas TC v8 overpredicted the number of cancers in this stratum. TC v7 and v8 had slightly lower discriminatory accuracy in women with two or more affected relatives (v7 AUC = 0.56, 95% CI = 0.51 to 0.61; v8 AUC = 0.56, 95% CI = 0.51 to 0.60) than in women with no family history (v7 AUC = 0.58, 95% CI = 0.54 to 0.61; v8 AUC = 0.59, 95% CI = 0.56 to 0.62) and women with one affected relative (v7 AUC = 0.62, 95% CI = 0.58 to 0.66; v8 AUC = 0.62, 95% CI = 0.58 to 0.66). In the subset of patients for whom the Claus model was applicable, the Claus model underpredicted the number of cancers (O/E = 1.69, 95% CI = 1.48 to 1.87), and the AUC was lower than in the other models (AUC = 0.59, 95% CI = 0.56 to 0.62). The positive predictive value was slightly higher for Claus than other models, but negative predictive value was lower.

BRCAPRO and Gail were well calibrated in both Ashkenazi Jewish and non-Ashkenazi Jewish women, whereas the TC models were well calibrated for the Ashkenazi Jewish women but overpredicted the number of cancers in non-Ashkenazi Jewish women (Table 5). All models had similar AUCs in Ashkenazi Jewish and non-Ashkenazi Jewish women, with AUCs ranging from 0.60 (95% CI = 0.55 to 0.67) to 0.66 (95% CI = 0.60 to 0.71) in the former and 0.61 (95% CI = 0.58 to 0.63) to 0.63 (95% CI = 0.61 to 0.65) in the latter. The TC models overpredicted

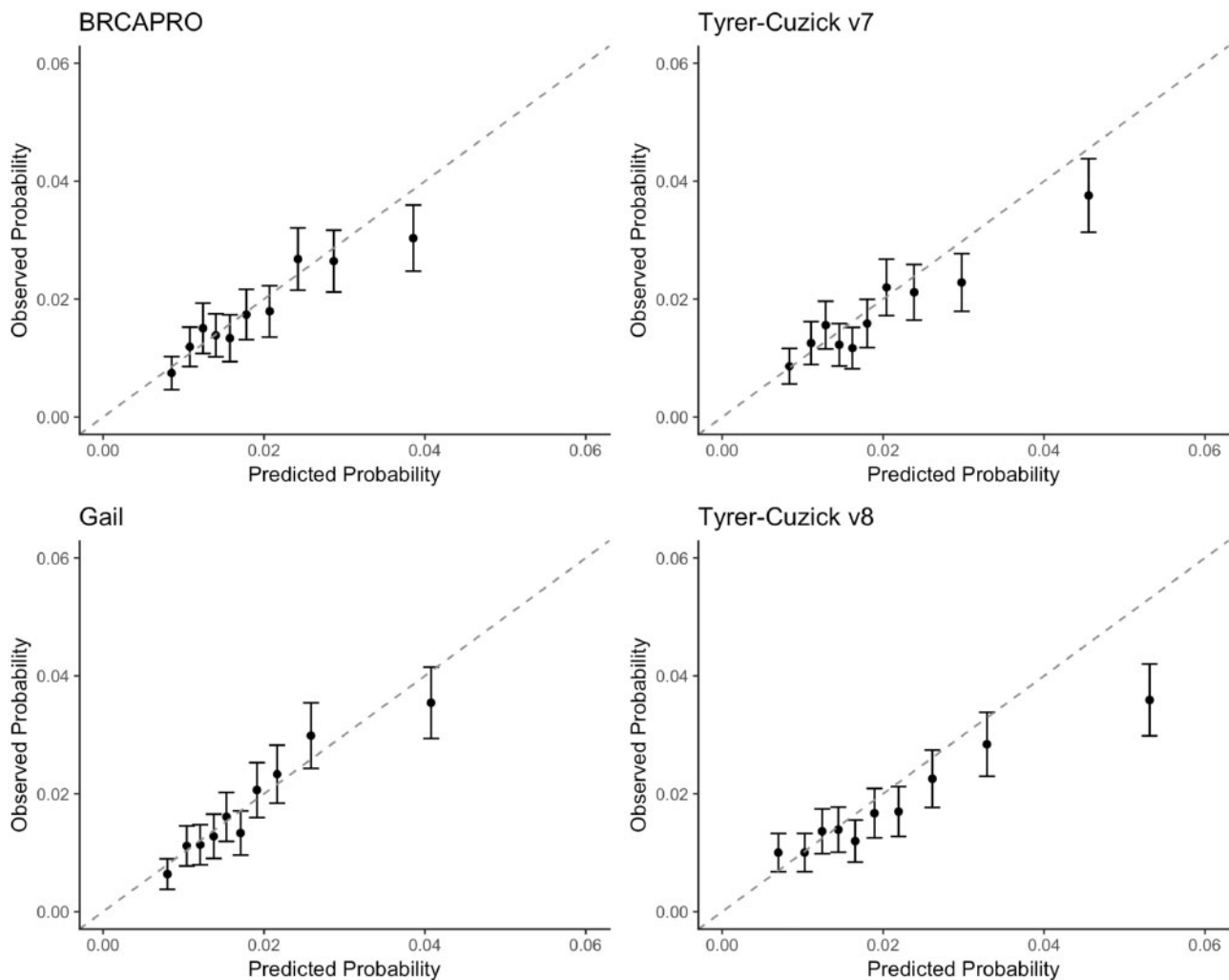


Figure 1. Calibration curves. Plots of the expected risk vs observed proportion of invasive breast cancer cases in each decile of risk. The error bars correspond to 95% Wald confidence intervals for the observed proportions. The dashed line is a 45-degree reference line where the observed risk equals the expected risk.

risk among women aged younger than 50 years, and BRCAPRO and TC v8 slightly overpredicted risk among women aged 50 years or older (Table 6). AUCs did not differ much by age group. The models performed best at predicting ER/PR+HER2–cancers and performed more poorly for HER2+ and triple-negative breast cancers; however, sample sizes were small for HER2+ ($N=42$) and triple-negative ($N=64$) breast cancers (Table 7).

For each model, density plots of the risk predictions by breast cancer status show that the distributions for affected and unaffected women are largely overlapping, with the distribution for affected women shifted slightly to the right (Supplementary Figure 2). For each pair of models, we calculated the overall correlation and correlation by cancer status (Supplementary Figure 2). The Gail and BCSC models were most highly correlated ($r=0.749$), and correlation was similar for both affected ($r=0.749$) and unaffected women ($r=0.747$). BRCAPRO and TC v8 were least correlated ($r=0.343$).

Discussion

To our knowledge, this is the first study to compare the predictive accuracy of five validated breast cancer risk models

(BRCAPRO, Gail, TC, BCSC, and Claus) in a large cohort of women undergoing mammography screening in the United States. Our results confirmed the comparable moderate discrimination of BRCAPRO, Gail, TC, and BCSC in predicting 6-year risk of cancer in the screening population. The Gail model had slightly higher discriminatory accuracy than BRCAPRO and TC did in the full study population, and the BCSC model had the highest discriminatory accuracy among women with available breast density information. The Claus model performed more poorly than the other models did. The predictive accuracy for HER2+ and triple-negative breast cancers was poorer than for ER/PR+HER2–cancers, though the sample sizes for the HER2+ and triple-negative categories were small. These results are useful to guide the choice of breast cancer risk-assessment tools for personalized breast cancer screening.

The measures of calibration and discrimination of the models in our study are consistent with the existing literature (9, 18–29). Some studies have compared multiple risk models in the same population, but most have focused on women with familial risk (18, 23, 28) or young women (27). In a prospective cohort of approximately 50 000 women undergoing mammography in the United Kingdom, the Gail and TC models were compared with a median follow-up of 3.2 years, and the TC model had a

Table 3. Performance of breast cancer risk-assessment models

Model	O/E (95% CI)	AUC (95% CI)	Sqrt(BS) (95% CI)	PPV (95% CI)	NPV (95% CI)
Overall performance (N = 35 921 647 invasive breast cancers)					
BRCAPRO	0.94 (0.88 to 1.02)	0.61 (0.59 to 0.63)	0.1328 (0.1284 to 0.1376)	0.026 (0.023 to 0.029)	0.987 (0.985 to 0.988)
Gail	0.98 (0.91 to 1.06)	0.64 (0.61 to 0.65)	0.1328 (0.1283 to 0.1376)	0.029 (0.026 to 0.032)	0.987 (0.986 to 0.988)
TC7	0.90 (0.84 to 0.96)	0.61 (0.59 to 0.63)	0.1329 (0.1284 to 0.1377)	0.026 (0.024 to 0.029)	0.987 (0.985 to 0.988)
TC8	0.84 (0.79 to 0.91)	0.62 (0.60 to 0.64)	0.1330 (0.1285 to 0.1377)	0.026 (0.024 to 0.028)	0.987 (0.986 to 0.989)
Among women with relevant data for BCSC model (N = 30 970 527 invasive breast cancers)*					
BRCAPRO	0.95 (0.87 to 1.03)	0.61 (0.58 to 0.63)	0.1292 (0.1241 to 0.1346)	0.025 (0.022 to 0.028)	0.987 (0.985 to 0.988)
Gail	0.95 (0.87 to 1.03)	0.63 (0.61 to 0.65)	0.1291 (0.1241 to 0.1344)	0.027 (0.024 to 0.030)	0.987 (0.986 to 0.989)
TC7	0.84 (0.77 to 0.91)	0.61 (0.59 to 0.64)	0.1292 (0.1242 to 0.1345)	0.025 (0.022 to 0.027)	0.987 (0.986 to 0.989)
TC8	0.78 (0.72 to 0.85)	0.63 (0.60 to 0.65)	0.1294 (0.1242 to 0.1345)	0.024 (0.022 to 0.027)	0.988 (0.987 to 0.990)
BCSC	0.97 (0.89 to 1.05)	0.64 (0.62 to 0.66)	0.1290 (0.1240 to 0.1344)	0.029 (0.026 to 0.032)	0.988 (0.987 to 0.989)

*Younger than age 75 years; breast density data available; no prior mastectomy or implants. AUC = area under the receiver operating characteristic curve; BCSC = Breast Cancer Surveillance Consortium; BS = Brier score; CI = confidence interval; NPV = negative predictive value; O/E = observed-to-expected ratio; PPV = positive predictive value; Sqrt = square root; TC = Tyrer-Cuzick.

slightly higher AUC than the Gail model did for 10-year risk (0.57 vs 0.55) (21). Another study from a general population cohort of over 60 000 women in the United Kingdom compared the 5-year performance of Gail and TC (29). In this cohort, Gail underestimated risk in women aged younger than 50 years and was well calibrated for women aged older than 50, whereas TC overestimated risk in both subgroups. TC had a higher AUC than Gail did in both subgroups.

In our study, the Gail and BCSC models had slightly better AUCs and calibration than the TC model did, although confidence intervals overlapped. This may be because the Gail and BCSC models were developed among US women, whereas much of the data used to develop TC were from non-US populations. Also, because the population comes from a mammography clinic rather than a high-risk clinic, the more comprehensive evaluation of family history used by the BRCAPRO, TC, and Claus models did not provide additional predictive accuracy for most women compared with the simple summary of family history used by Gail and BCSC models. We had limited data on unaffected relatives, so risk estimates for BRCAPRO and TC may be different than if a full pedigree was available. Prior work suggests that this can lead to overestimation of risk if no adjustment is made for missing unaffected relatives, but this has little impact on discrimination (30). By default, BRCAPRO imputes the number and ages of unaffected relatives in families with none reported, potentially reducing the effect on calibration. Simpler family history may suffice for a screening population, whereas the more extensive family history likely has greater value in the context of genetic counseling for women with extensive family histories.

Though the AUC values were comparable across most models, the correlation of risk estimates was only moderate. This highlights that various models identify different subgroups of women as high risk, and an individual may be identified as high risk by one model and not another. This is not surprising because the models use different risk factors and were developed on different training populations. Gail and BCSC consider mainly reproductive and hormonal risk factors; Claus and BRCAPRO rely exclusively or almost exclusively on detailed family history information; and TC combines reproductive and hormonal risk factors with detailed family history information. Future work

should evaluate whether combinations of models that are not highly correlated may yield better estimation of risk.

Our results suggest better predictive accuracy for ER/PR+ cancers than HER2+ and triple-negative breast cancers, though our sample sizes were small. A study of the Women's Health Initiative showed that the Gail model predicted ER+ breast cancers, but not ER- breast cancers (24). Because ER+ cancers are far more common than ER- cancers, the models are trained to learn risk factors related primarily to ER+ disease. It is increasingly clear that both etiology and incidence patterns vary by tumor subtype. Leveraging these differences may allow recalibration of existing models or creation of new models that estimate risk for different breast cancer subtypes to increase predictive accuracy.

Our study is the largest to simultaneously assess the performance of the BRCAPRO, Gail, TC, BCSC, and Claus models in a screening setting in the United States. Virtually every woman who receives a mammogram at Newton-Wellesley Hospital first completes a comprehensive breast cancer risk assessment, which provided a large, unselected clinical population with detailed risk information on which to evaluate the models.

Despite these strengths, several limitations must be considered. Missing information on unaffected relatives may have affected risk estimates for BRCAPRO and TC, especially the calibration of these models. Another widely used family history-based breast cancer risk prediction model is BOADICEA (31), but we did not include BOADICEA in our analyses because of the lack of easily accessible software for offline use. Our analysis highlights the inherent difficulty in comparing accuracy across models, which vary in their exclusion criteria and handling of missing data. We attempted to compare models across similar populations using measures of both discrimination and calibration. Additionally, the screening population was relatively young, with 42.4% of women in their 40s. Results may not generalize to older women. However, women in their 40s may derive the most benefit from risk assessment, so having a large sample in this key age group could also be viewed as a strength. Furthermore, we assessed 6-year risk, but additional studies are needed to evaluate model performance for longer-term risk estimates, particularly because breast magnetic resonance

Table 4. Model performance by family history

Model	O/E (95% CI)	AUC (95% CI)	Sqrt(BS) (95% CI)	PPV (95% CI)	NPV (95% CI)
No family history of breast or ovarian cancer in first- or second-degree relatives (N = 22 927, 64%, 343 invasive breast cancers)					
BRCAPRO	0.82 (0.74 to 0.91)	0.59 (0.57 to 0.62)	0.1214 (0.1155 to 0.1281)	0.022 (0.019 to 0.026)	0.989 (0.987 to 0.990)
Gail	0.95 (0.86 to 1.06)	0.61 (0.58 to 0.64)	0.1213 (0.1154 to 0.1280)	0.024 (0.020 to 0.028)	0.988 (0.986 to 0.989)
TC7	0.99 (0.89 to 1.10)	0.58 (0.54 to 0.61)	0.1214 (0.1155 to 0.1281)	0.023 (0.018 to 0.027)	0.986 (0.985 to 0.988)
TC8	0.93 (0.84 to 1.03)	0.59 (0.56 to 0.62)	0.1213 (0.1155 to 0.1281)	0.022 (0.019 to 0.026)	0.987 (0.986 to 0.989)
One first- or second-degree family member with breast or ovarian cancer (N = 9391, 26%, 186 invasive breast cancers)					
BRCAPRO	1.02 (0.88 to 1.16)	0.61 (0.57 to 0.64)	0.1392 (0.1288 to 0.1478)	0.028 (0.023 to 0.034)	0.985 (0.982 to 0.988)
Gail	0.93 (0.80 to 1.06)	0.63 (0.58 to 0.66)	0.1390 (0.1286 to 0.1476)	0.028 (0.023 to 0.034)	0.987 (0.984 to 0.990)
TC7	0.77 (0.66 to 0.87)	0.62 (0.58 to 0.66)	0.1392 (0.1289 to 0.1477)	0.024 (0.021 to 0.028)	0.989 (0.985 to 0.992)
TC8	0.73 (0.62 to 0.82)	0.62 (0.58 to 0.66)	0.1394 (0.1293 to 0.1480)	0.024 (0.021 to 0.028)	0.989 (0.985 to 0.992)
Two or more first- or second-degree family members with breast or ovarian cancer (N = 3603, 10%, 118 invasive breast cancers)					
BRCAPRO	1.40 (1.20 to 1.65)	0.60 (0.55 to 0.65)	0.1779 (0.1647 to 0.1935)	0.042 (0.032 to 0.051)	0.977 (0.970 to 0.983)
Gail	1.19 (1.03 to 1.41)	0.61 (0.57 to 0.65)	0.1784 (0.1658 to 0.1937)	0.043 (0.035 to 0.052)	0.981 (0.974 to 0.987)
TC7	0.89 (0.77 to 1.06)	0.56 (0.51 to 0.61)	0.1783 (0.1657 to 0.1935)	0.035 (0.029 to 0.041)	0.982 (0.969 to 0.993)
TC8	0.83 (0.71 to 0.98)	0.56 (0.51 to 0.60)	0.1787 (0.1662 to 0.1939)	0.035 (0.029 to 0.041)	0.979 (0.967 to 0.990)
Among women with relevant data for Claus model (N = 11 873, 278 invasive breast cancers; 52 ductal carcinoma in situ)*					
BRCAPRO	1.18 (1.03 to 1.31)	0.62 (0.59 to 0.65)	0.1510 (0.1414 to 0.1590)	0.033 (0.027 to 0.038)	0.983 (0.980 to 0.986)
Gail	1.02 (0.89 to 1.13)	0.64 (0.60 to 0.67)	0.1510 (0.1415 to 0.1590)	0.033 (0.028 to 0.037)	0.985 (0.982 to 0.989)
TC7	0.79 (0.69 to 0.88)	0.63 (0.60 to 0.66)	0.1511 (0.1418 to 0.1590)	0.028 (0.024 to 0.031)	0.991 (0.987 to 0.994)
TC8	0.74 (0.64 to 0.82)	0.63 (0.60 to 0.66)	0.1514 (0.1422 to 0.1591)	0.028 (0.024 to 0.031)	0.989 (0.986 to 0.992)
Claust†	1.69 (1.48 to 1.87)	0.59 (0.56 to 0.62)	0.1646 (0.1543 to 0.1731)	0.039 (0.031 to 0.046)	0.975 (0.972 to 0.978)

*Aged younger than 79 years with at least one first- or second-degree relative with breast cancer or at least one first-degree relative with ovarian cancer.

†Validated against diagnosis of ductal carcinoma in situ or invasive breast cancer. AUC = area under the receiver operating characteristic curve; BS = Brier score; CI = confidence interval; NPV = negative predictive value; O/E = observed-to-expected ratio; PPV = positive predictive value; Sqrt = square root; TC = Tyrer-Cuzick.

Table 5. Model performance by Ashkenazi Jewish ancestry*

Model	O/E (95% CI)	AUC (95% CI)	Sqrt(BS) (95% CI)	PPV (95% CI)	NPV (95% CI)
Ashkenazi Jewish women (N = 5 165 114 invasive breast cancers)					
BRCAPRO	0.95 (0.79 to 1.12)	0.61 (0.56 to 0.67)	0.1466 (0.1343 to 0.1586)	0.029 (0.023 to 0.036)	0.985 (0.981 to 0.990)
Gail	1.13 (0.95 to 1.34)	0.66 (0.60 to 0.71)	0.1465 (0.1338 to 0.1586)	0.037 (0.028 to 0.046)	0.986 (0.982 to 0.990)
TC7	0.98 (0.81 to 1.16)	0.60 (0.55 to 0.67)	0.1468 (0.1344 to 0.1587)	0.032 (0.024 to 0.039)	0.985 (0.980 to 0.989)
TC8	0.91 (0.75 to 1.08)	0.63 (0.57 to 0.69)	0.1468 (0.1345 to 0.1587)	0.031 (0.024 to 0.039)	0.987 (0.982 to 0.991)
Non-Ashkenazi Jewish women or unknown ancestry (N = 30 756, N = 533 invasive breast cancers)					
BRCAPRO	0.94 (0.87 to 1.01)	0.61 (0.58 to 0.63)	0.1304 (0.1259 to 0.1350)	0.026 (0.023 to 0.029)	0.987 (0.986 to 0.988)
Gail	0.95 (0.88 to 1.02)	0.63 (0.61 to 0.65)	0.1304 (0.1258 to 0.1350)	0.028 (0.025 to 0.032)	0.987 (0.986 to 0.989)
TC7	0.88 (0.82 to 0.94)	0.61 (0.59 to 0.64)	0.1304 (0.1259 to 0.1351)	0.025 (0.022 to 0.028)	0.987 (0.985 to 0.988)
TC8	0.83 (0.77 to 0.89)	0.62 (0.59 to 0.64)	0.1305 (0.1260 to 0.1352)	0.025 (0.022 to 0.028)	0.987 (0.986 to 0.989)

*AUC = area under the receiver operating characteristic curve; BS = Brier score; CI = confidence interval; NPV = negative predictive value; O/E = observed-to-expected ratio; PPV = positive predictive value; Sqrt = square root; TC = Tyrer-Cuzick.

Table 6. Model performance by age*

Model	O/E (95% CI)	AUC (95% CI)	Sqrt(BS) (95% CI)	PPV (95% CI)	NPV (95% CI)
Aged younger than 50 y (N = 15 230 193 invasive breast cancers)					
BRCAPRO	1.03 (0.92 to 1.18)	0.58 (0.54 to 0.62)	0.1118 (0.1054 to 0.1198)	0.026 (0.014 to 0.039)	0.988 (0.986 to 0.989)
Gail	0.95 (0.84 to 1.09)	0.62 (0.57 to 0.66)	0.1117 (0.1052 to 0.1196)	0.029 (0.022 to 0.037)	0.989 (0.987 to 0.991)
TC7	0.79 (0.70 to 0.91)	0.60 (0.56 to 0.63)	0.1119 (0.1056 to 0.1198)	0.021 (0.017 to 0.026)	0.989 (0.988 to 0.991)
TC8	0.76 (0.67 to 0.87)	0.60 (0.56 to 0.64)	0.1119 (0.1057 to 0.1199)	0.02 (0.016 to 0.023)	0.99 (0.987 to 0.991)
Aged 50 y or older (N = 20 691, N = 454 invasive breast cancers)					
BRCAPRO	0.91 (0.84 to 0.99)	0.58 (0.55 to 0.61)	0.1464 (0.1405 to 0.1526)	0.026 (0.024 to 0.029)	0.985 (0.982 to 0.988)
Gail	0.99 (0.91 to 1.08)	0.61 (0.58 to 0.63)	0.1464 (0.1405 to 0.1527)	0.029 (0.027 to 0.032)	0.984 (0.982 to 0.987)
TC7	0.95 (0.88 to 1.04)	0.59 (0.57 to 0.61)	0.1464 (0.1404 to 0.1526)	0.028 (0.025 to 0.031)	0.983 (0.981 to 0.986)
TC8	0.89 (0.81 to 0.97)	0.60 (0.57 to 0.63)	0.1466 (0.1406 to 0.1528)	0.028 (0.025 to 0.031)	0.985 (0.983 to 0.987)

*AUC = area under the receiver operating characteristic curve; BS = Brier score; CI = confidence interval; NPV = negative predictive value; O/E = observed-to-expected ratio; PPV = positive predictive value; Sqrt = square root; TC = Tyrer-Cuzick.

Table 7. Performance of breast cancer risk-assessment models by tumor subtype*

Model	AUC (95% CI)	Sqrt(BS) (95% CI)	PPV (95% CI)	NPV (95% CI)
Estrogen receptor/ progesterone receptor+ HER2- (395 invasive breast cancers)				
BRCAPRO	0.64 (0.61 to 0.67)	0.1051 (0.1005 to 0.1094)	0.018 (0.015 to 0.020)	0.993 (0.991 to 0.994)
Gail	0.65 (0.63 to 0.68)	0.1050 (0.1004 to 0.1094)	0.019 (0.016 to 0.022)	0.992 (0.991 to 0.993)
TC7	0.62 (0.59 to 0.64)	0.1053 (0.1007 to 0.1097)	0.016 (0.014 to 0.018)	0.992 (0.991 to 0.993)
TC8	0.63 (0.60 to 0.66)	0.1056 (0.1010 to 0.1099)	0.017 (0.015 to 0.019)	0.993 (0.991 to 0.994)
HER2+ (42 invasive breast cancers)				
BRCAPRO	0.51 (0.41 to 0.59)	0.0399 (0.0356 to 0.0447)	0.001 (0.001 to 0.002)	0.999 (0.998 to 0.999)
Gail	0.50 (0.39 to 0.59)	0.0398 (0.0355 to 0.0448)	0.001 (0.001 to 0.002)	0.999 (0.998 to 0.999)
TC7	0.55 (0.44 to 0.64)	0.0408 (0.0365 to 0.0456)	0.002 (0.001 to 0.003)	0.999 (0.999 to 0.999)
TC8	0.57 (0.47 to 0.66)	0.0421 (0.0381 to 0.0468)	0.002 (0.001 to 0.002)	0.999 (0.999 to 0.999)
Triple-negative (64 invasive breast cancers)				
BRCAPRO	0.55 (0.49 to 0.63)	0.0467 (0.0419 to 0.0507)	0.002 (0.002 to 0.003)	0.999 (0.998 to 0.999)
Gail	0.60 (0.54 to 0.68)	0.0466 (0.0417 to 0.0506)	0.003 (0.002 to 0.004)	0.999 (0.998 to 0.999)
TC7	0.55 (0.49 to 0.63)	0.0475 (0.0427 to 0.0515)	0.002 (0.001 to 0.003)	0.998 (0.998 to 0.999)
TC8	0.52 (0.45 to 0.61)	0.0487 (0.0440 to 0.0526)	0.002 (0.001 to 0.003)	0.998 (0.998 to 0.999)

*AUC = area under the receiver operating characteristic curve; BS = Brier score; CI = confidence interval; HER = human epidermal growth factor receptor; NPV = negative predictive value; PPV = positive predictive value; Sqrt = square root; TC = Tyrer-Cuzick.

imaging recommendations are based on lifetime risk. Finally, our study population was predominantly white, and additional studies in nonwhite populations are warranted.

Our results are reassuring in that the BRCAPRO, Gail, TC, and BCSC models have similar calibration and predictive accuracy in a mammography screening population. However, model discrimination is only modest, and this modest discrimination in conjunction with the time and difficulty of obtaining risk factors have led to limited adoption of breast cancer risk-assessment models outside of high-risk clinics. Employing technology that can electronically obtain risk factors and automatically calculate risk will be key to enabling adoption of personalized screening approaches. Additionally, improving risk models by adding risk factors or by differentiating risk by tumor subtypes may further improve risk prediction and thereby lead to increased use of breast cancer risk-assessment models in routine clinical care.

Funding

This work was supported by the American Cancer Society (131052-MRSG-17-144-01-CCE to AMM); the Susan G. Komen

Foundation (CCR17480662 to AMM); the National Institutes of Health and National Cancer Institute (5P30 CA006516-50 to GP); a Natural Sciences and Engineering Research Council of Canada PGS-D Scholarship (PGSD3502362-2017 to ZG); and the Research Scientist Development Fund at the Dana-Farber Cancer Institute to DB.

Notes

Affiliations of authors: Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA (AMM); Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA (MW); Department of Biostatistics, Harvard University T H Chan School of Public Health, Boston, MA (ZG, GP, DB); Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA (ZG, GP, DB); Division of Surgical Oncology, Massachusetts General Hospital, Boston, MA (MEG, ZD, SBC, KSH); Department of Radiology, Massachusetts General Hospital, Boston, MA (DAS); Istanbul School of Medicine,

Istanbul University, Istanbul, Turkey (AA); Department of Radiology, Newton-Wellesley Hospital, Newton, MA (AS).

The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

KSH receives Honoraria from Hologic (surgical implant for radiation planning with breast conservation and wire free breast biopsy) and 23andMe and is a founder of and has a financial interest in CRA Health (formerly Hughes RiskApps). His interests were reviewed and are managed by Massachusetts General Hospital and Partners Health Care in accordance with their conflict of interest policies. DB and GP co-lead the BayesMendel laboratory, which licenses software for the computation of the BRCAPRO and other models. These authors do not derive any personal income from these licenses. All revenues are assigned to the lab for software maintenance and upgrades. GP holds equity in CRA Health. The other authors have no disclosures.

References

- American Cancer Society. *Breast Cancer Facts & Figures 2017–2018*. Atlanta, GA: American Cancer Society, Inc.; 2017.
- Cintolo-Gonzalez JA, Braun D, Blackford AL, et al. Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. *Breast Cancer Res Treat*. 2017;164(2):263–284.
- Barnard ME, Boeke CE, Tamimi RM. Established breast cancer risk factors and risk of intrinsic tumor subtypes. *Biochim Biophys Acta*. 2015;1856(1):73–85.
- Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879–1886.
- Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst*. 2007;99(23):1782–1792.
- Matsuno RK, Costantino JP, Ziegler RG, et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. *J Natl Cancer Inst*. 2011;103(12):951–961.
- Banegas MP, John EM, Slattery ML, et al. Projecting individualized absolute invasive breast cancer risk in US Hispanic women. *J Natl Cancer Inst*. 2017;109(2):djw215.
- Berry DA, Parmigiani G, Sanchez J, et al. Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *J Natl Cancer Inst*. 1997;89(3):227–238.
- 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):337–347.
- Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. *Am J Hum Genet*. 1991;48(2):232–242.
- Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer*. 1994;73(3):643–651.
- Claus EB, Risch N, Thompson WD. The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. *Breast Cancer Res Treat*. 1993;28(2):115–120.
- Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med*. 2004;23(7):1111–1130.
- Parmigiani G, Chen S, Wang W, et al. *Bayes Mendel: Determining Carrier Probabilities for Cancer Susceptibility Genes*; 2018. <https://bcf.dfci.harvard.edu/bayesmendel/index.php>. Accessed August 7, 2019.
- Zhang F. BCRA: Breast Cancer Risk Assessment; 2018. <https://cran.r-project.org/package=BCRA>. Accessed August 7, 2019.
- IBIS Breast Cancer Risk Evaluation Tool; 2017. <http://www.ems-trials.org/risk-evaluator/>. Accessed August 7, 2019.
- NCI-Funded Breast Cancer Surveillance Consortium (P01 CA154292 and HHSN261201100031C). <https://tools.bccsc-scc.org/BC5yearRisk>. Accessed August 7, 2019.
- Amir E, Evans DG, Shenton A, et al. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *J Med Genet*. 2003;40(11):807–814.
- Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Risk prediction models of breast cancer: a systematic review of model performances. *Breast Cancer Res Treat*. 2012;133(1):1–10.
- Wang X, Huang Y, Li L, et al. Assessment of performance of the Gail model for predicting breast cancer risk: a systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Res*. 2018;20(1):18.
- Brentnall AR, Harkness EF, Astley SM, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res*. 2015;17(1):147.
- Brentnall AR, Cuzick J, Buist DSM, et al. Long-term accuracy of breast cancer risk assessment combining classic risk factors and breast density. *JAMA Oncol*. 2018;4(9):e180174.
- Quante AS, Whittemore AS, Shriver T, et al. Breast cancer risk assessment across the risk continuum: genetic and non genetic risk factors contributing to differential model performance. *Breast Cancer Res*. 2012;14(6):R144.
- Chlebowski RT, Anderson GL, Lane DS, et al. Predicting risk of breast cancer in postmenopausal women by hormone receptor status. *J Natl Cancer Inst*. 2007;99(22):1695–1705.
- McTiernan A, Kuniyuki A, Yasui Y, et al. Comparisons of two breast cancer risk estimates in women with a family history of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2001;10(4):333–338.
- Rockhill B, Spiegelman D, Byrne C, et al. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst*. 2001;93(5):358–366.
- Dite GS, MacInnis RJ, Bickerstaffe A, et al. Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women aged under 50 years: Australian Breast Cancer Family Registry. *Cancer Epidemiol Biomarkers Prev*. 2016;25(2):359–365.
- Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. *Lancet Oncol*. 2019;20(4):504–517.
- Choudhury PP, Wilcox A, Brook M, et al. Comparative validation of breast cancer risk prediction models and projections for future risk stratification. *J Natl Cancer Inst*. 2019; doi: 10.1093/jnci/djz113.
- Biswas S, Atienza P, Chipman J, et al. Simplifying clinical use of the genetic risk prediction model BRCAPRO. *Breast Cancer Res Treat*. 2013;139(2):571–579.
- Antoniou AC, Pharoah PP, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer*. 2004;91(8):1580.