

ARTICLE

Comparative Validation of Breast Cancer Risk Prediction Models and Projections for Future Risk Stratification

Parichoy Pal Choudhury,* Amber N. Wilcox,* Mark N. Brook, Yan Zhang, Thomas Ahearn, Nick Orr, Penny Coulson, Minouk J. Schoemaker, Michael E. Jones, Mitchell H. Gail, Anthony J. Swerdlow, Nilanjan Chatterjee,* Montserrat Garcia-Closas*

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

*These authors contributed equally to this work.

Correspondence to: Nilanjan Chatterjee, PhD, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Room E3612, BSPH, Baltimore, MD 21205 (e-mail: nchatte2@jhu.edu).

Abstract

Background: External validation of risk models is critical for risk-stratified breast cancer prevention. We used the Individualized Coherent Absolute Risk Estimation (iCARE) as a flexible tool for risk model development and comparative model validation and to make projections for population risk stratification.

Methods: Performance of two recently developed models, one based on the Breast and Prostate Cancer Cohort Consortium analysis (iCARE-BPC3) and another based on a literature review (iCARE-Lit), were compared with two established models (Breast Cancer Risk Assessment Tool and International Breast Cancer Intervention Study Model) based on classical risk factors in a UK-based cohort of 64 874 white non-Hispanic women (863 patients) age 35–74 years. Risk projections in a target population of US white non-Hispanic women age 50–70 years assessed potential improvements in risk stratification by adding mammographic breast density (MD) and polygenic risk score (PRS).

Results: The best calibrated models were iCARE-Lit (expected to observed number of cases [E/O] = 0.98, 95% confidence interval [CI] = 0.87 to 1.11) for women younger than 50 years, and iCARE-BPC3 (E/O = 1.00, 95% CI = 0.93 to 1.09) for women 50 years or older. Risk projections using iCARE-BPC3 indicated classical risk factors can identify approximately 500 000 women at moderate to high risk (>3% 5-year risk) in the target population. Addition of MD and a 313-variant PRS is expected to increase this number to approximately 3.5 million women, and among them, approximately 153 000 are expected to develop invasive breast cancer within 5 years.

Conclusions: iCARE models based on classical risk factors perform similarly to or better than BCRAT or IBIS in white non-Hispanic women. Addition of MD and PRS can lead to substantial improvements in risk stratification. However, these integrated models require independent prospective validation before broad clinical applications.

Breast cancer risk prediction models are used in clinical and research settings to identify women at elevated risk of disease who could benefit from preventive therapies and enhanced screening or be eligible to participate in prevention trials. Continuing updates of risk models incorporating additional risk factors will potentially improve our ability to identify such women (1), as well as women at low risk of disease.

Independent prospective validation of models is critical to determine their accuracy of prediction, robustness, and potential for clinical application. Breast Cancer Risk Assessment Tool (BCRAT) and International Breast Cancer Intervention Study Model (IBIS) are established models that are currently used for clinical and research applications (2). BCRAT includes classical risk factors and has been extensively evaluated generally

Received: October 9, 2018; Revised: January 31, 2019; Accepted: May 29, 2019

Published by Oxford University Press 2019. This work is written by US Government employees and is in the public domain in the US.

showing good calibration but low risk discrimination (2–4). IBIS, which includes a more comprehensive set of classical risk factors and extensive family history, performed better than BCRAT in average- to high-risk populations (5,6). Although addition of mammographic breast density (MD) (7–11) or polygenic risk scores (PRS) (12–20) can lead to improved risk stratification, prospective evaluation of the accuracy of absolute risk predictions from models incorporating PRS is limited (21).

Risk prediction models should be dynamic and flexible in their ability to incorporate additional risk factors and context-specific incidence rates. However, developing and validating a comprehensive model is challenging because of all the relevant risk factors not being typically measured in a single study and requiring novel methods for data integration from multiple epidemiologic studies (22–24). Our recently developed Individualized Coherent Absolute Risk Estimation (iCARE) software implements a flexible approach to build absolute risk models for a population combining information on relative risk estimates, age-specific incidence and mortality rates, and risk factor distributions from multiple data sources (25–27). It includes advanced features to account for missing risk factors using internal imputation and a validation component to facilitate comparative model evaluation across multiple cohorts using uniform methodology (27).

We previously used iCARE to develop a breast cancer risk model using relative risks from a multivariable regression based on eight prospective cohorts of women age 50 years or older in the Breast and Prostate Cancer Cohort Consortium (iCARE-BPC3) (25). Here, we develop an updated version of the synthetic model, described in Garcia-Closas et al. (1), using relative risks from published literature (iCARE-Lit). A literature-based model, though requiring more assumptions, can include comprehensive sets of risk factors that may not all be measured in one study.

The current study aims to compare the performances of the iCARE models, BCRAT, and IBIS based on classical risk factors (ie, questionnaire-based risk factors like menstrual, reproductive, hormonal, and lifestyle risk factors) in the UK-based Generations Study (GS). Additionally, the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, used to develop iCARE-BPC3, provided further evaluation of the other models. Risk projections in a target population were estimated based on classical risk factors and after addition of MD (28) and PRS.

Materials and Methods

Study Populations

Primary analyses were performed in a population of 113 211 women age 16–102 years at enrollment (2003–2012) from the UK-based GS. Further validation of the iCARE-Lit model was performed in 78 214 women age 50–75 years at enrollment (1993–2001) from the US-based PLCO. Exclusion criteria included history of breast cancer, nonwhite or unknown ethnicity, no genetic consent or DNA source, entry age younger than 35 or older than 75 years, presence of first- or second-degree relative in study (GS only), and individuals with an unconfirmable report of breast cancer (PLCO only). The final analytic samples from the GS and PLCO were 64 874 (863 cases within 5 years) and 48 279 (1008 cases within 5 years), respectively (Supplementary Figure 1, available online). As PLCO was used for the development of iCARE-BPC3 (24), it was used only for validating other models.

Supplementary Table 1 (available online) shows the risk factor distributions in both cohorts.

Breast Cancer Risk Model Validation and Risk Projection

Supplementary Tables 2–4 (available online) provide detailed descriptions of the iCARE-based models, BCRAT and IBIS. All models incorporate information on marginal disease incidence rates (Supplementary Figure 2, A and B, available online) and account for competing mortality using mortality rates, both available from population-based registries (27,29,30). The incidence rates were used to calibrate the average predicted risk to the national breast cancer risk (29,30). iCARE implements this step using an additional individual-level reference dataset of risk factors representing the underlying population.

For evaluating calibration, we categorized individuals based on deciles of both of the 5-year absolute risks that incorporate the variation of age and the relative risk score (ie, sum of log relative risks multiplied by risk factors) that does not include age. The predicted and observed risks across risk categories were compared using expected-to-observed (E/O) ratio, calibration slope, and intercept. Model discrimination was assessed using area under the curve (AUC) statistics based both on 5-year absolute risk and the relative risk score (Supplementary Methods, available online).

Risk projections of invasive breast cancer were estimated among US white non-Hispanic women age 50–70 years using the best calibrated model based on classical risk factors in that group. We also evaluated the net benefit (31–33) of this model for high-risk decisions in that population (Supplementary Methods, available online). We explored potential improvements in risk stratification and net benefit with addition of PRS and MD. Apart from the PRS constructed using 313 single-nucleotide polymorphisms (SNPs) (Supplementary Table 5, available online) (34–36), we considered an “improved” PRS incorporating the fraction of additional heritability attributable to common variants, and a “best” PRS incorporating all the common variant heritability (34,36–38). Theoretical AUC was computed using a normal approximation of the relative risk scores for different combinations: classical risk factors only, PRS only, MD only, and a combined model with all risk factors (26,38,39). Moreover, we considered two high-risk thresholds: 3% corresponding to US Preventive Services Task Force recommendation for risk-lowering drugs and 6% used by the WISDOM trial as a cutoff for very high risk (40,41) and two low risk thresholds: 0.6% and 1.3%, which are average 5-year risks of US women age 40 and 50 years, respectively. We estimated numbers of women and future cases identified at the extremes of the risk distribution based on the above thresholds (Supplementary Methods, available online) (25,27).

Results

Breast Cancer Risk Model Validation

Among women younger than 50 years, all models showed good calibration of relative risk (Figure 1, Table 1; Supplementary Figure 3A, available online). Absolute risk was best calibrated for iCARE-Lit (E/O = 0.98, 96% CI = 0.87 to 0.11), with a discrimination of AUC = 65.4 (95% CI = 62.1 to 68.7) (Table 1). BCRAT tended to underestimate (E/O = 0.85, 95% CI = 0.75 to 0.95) and IBIS to overestimate (E/O = 1.14, 95% CI = 1.01 to 1.29) absolute risk.

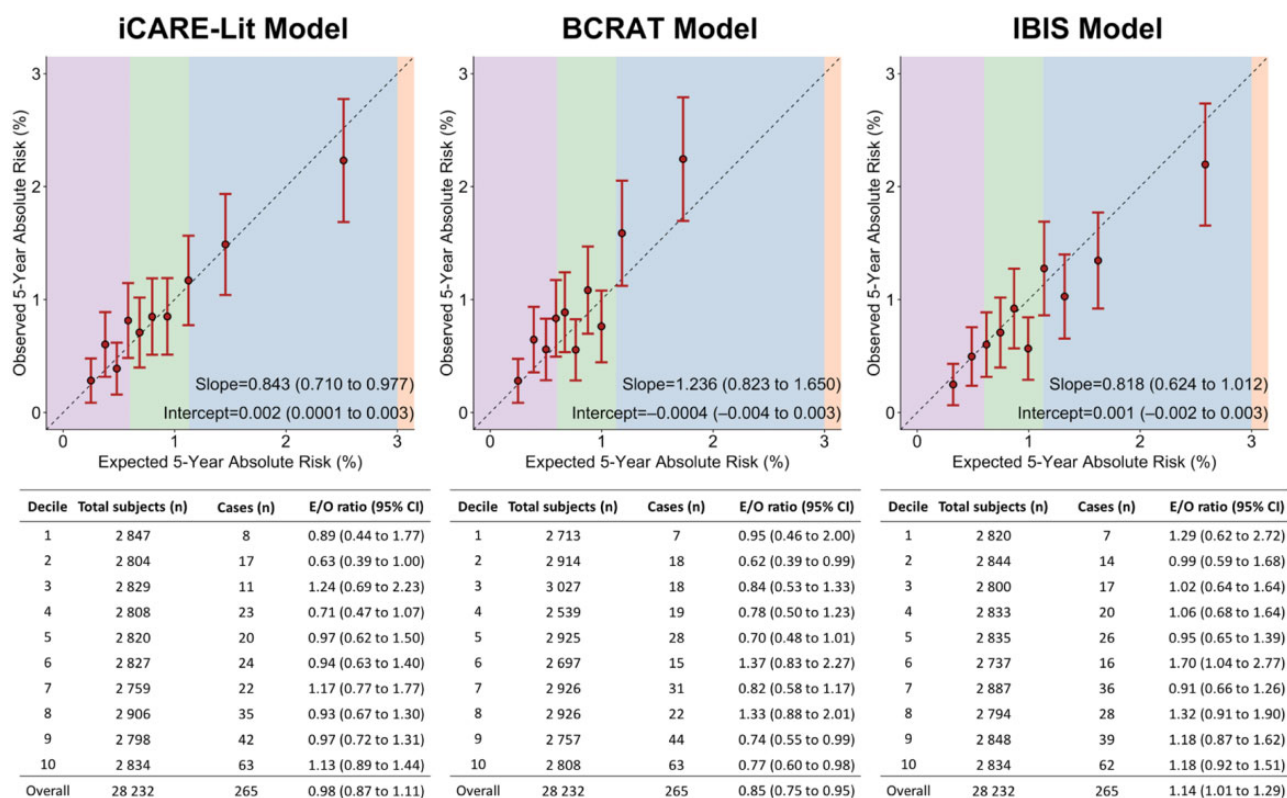


Figure 1. Absolute risk calibration of breast cancer risk prediction models in the GS cohort among women younger than 50 years. The risk categories are based on absolute risk. The backgrounds of the plots are shaded to indicate the absolute risk threshold categories ($\leq 0.6\%$ in pink, $>0.6\%$ to $\leq 1.13\%$ in green, $>1.13\%$ to $\geq 3\%$ in blue, and $>3\%$ in orange). The 0.6% and 1.13% thresholds correspond to the average 5-year risk for US women age 40 years and 50 years, respectively. The 3% threshold is used by the US Preventive Services Task Force for recommending risk-lowering drugs, and 6% is used by the WISDOM trial as a threshold for very high risk. BCRAT = Breast Cancer Risk Assessment Tool; CI = confidence interval; E/O = expected-to-observed; GS = Generations Study; IBIS = International Breast Cancer Intervention Study; iCARE-Lit = Individualized Coherent Absolute Risk Estimation model based on literature review.

Table 1. Ratios of expected-to-observed 5-year absolute risk for the breast cancer risk prediction models validated using the GS*

Age group, years (number of cases, number of noncases)	Model	AUC (95% CI)	Overall				Top risk decile		
			O% (95% CI)	E%	E/O ratio (95% CI)	O% (95% CI)	E%	E/O ratio (95% CI)	
<50 (265 cases, 27 967 noncases)	iCARE-Lit	65.4 (62.1 to 68.7)	0.94 (0.83 to 1.05)	0.92	0.98 (0.87 to 1.11)	2.22 (1.68 to 2.77)	2.51	1.13 (0.89 to 1.44)	
	BCRAT	64.0 (60.6 to 67.4)		0.79	0.85 (0.75 to 0.95)	2.24 (1.70 to 2.79)	1.73	0.77 (0.60 to 0.98)	
	IBIS	64.6 (61.3 to 67.9)		1.07	1.14 (1.01 to 1.29)	2.19 (1.65 to 2.73)	2.58	1.18 (0.92 to 1.51)	
≥ 50 (598 cases, 36 044 noncases)	iCARE-Lit	62.2 (60.0 to 64.5)	1.63 (1.50 to 1.76)	1.84	1.13 (1.04 to 1.22)	3.20 (2.62 to 3.77)	3.91	1.22 (1.02 to 1.46)	
	iCARE-BPC3	60.2 (58.0 to 62.4)		1.64	1.00 (0.93 to 1.09)	2.63 (2.12 to 3.15)	2.85	1.08 (0.89 to 1.32)	
	BCRAT	58.2 (55.8 to 60.5)		1.56	0.95 (0.88 to 1.03)	2.73 (2.20 to 3.26)	3.27	1.20 (0.99 to 1.46)	
	IBIS	61.4 (59.2 to 63.6)		1.85	1.13 (1.05 to 1.23)	3.07 (2.51 to 3.63)	3.98	1.30 (1.08 to 1.56)	

*The AUCs reported in Table 1 are defined based on absolute risk and incorporate the variation due to age. The AUCs (95% CI) based on the relative risk score, which do not incorporate variation of age, are as follows: for women younger than 50 years, iCARE-Lit: 58.8 (95% CI = 55.3 to 62.3); BCRAT: 54.6 (95% CI = 50.9 to 58.4); IBIS: 57.0 (95% CI = 53.4 to 60.6); and for women 50 years or older, iCARE-Lit: 60.3 (95% CI = 58.0 to 62.6); iCARE-BPC3: 57.7 (95% CI = 55.4 to 60.0); BCRAT: 52.2 (95% CI = 49.6 to 54.7); IBIS: 60.2 (95% CI = 57.9 to 62.5). AUC = area under the curve; GS = Generations Study; CI = confidence interval; BCRAT = Breast Cancer Risk Assessment Tool; E = expected absolute risk; E/O = expected-to-observed; iCARE-BPC3 = Individualized Coherent Absolute Risk Estimation model based on Breast and Prostate Cancer Cohort Consortium; iCARE-Lit = iCARE model based on literature review; IBIS = International Breast Cancer Intervention Study; O = observed absolute risk.

Among women age 50 years or older, iCARE-BPC3 showed good calibration of absolute and relative risk (E/O = 1.00, 95% CI = 0.93 to 1.09), with AUC = 60.2, 95% CI = 58.0 to 62.4. iCARE-Lit showed good calibration of relative risk but overestimation of absolute risk (E/O = 1.13, 95% CI = 1.04 to 1.22) (Table 1, Figure 2; Supplementary Figure 4A, available online). BCRAT and IBIS showed miscalibration both of absolute and relative risk. BCRAT tended to show underestimation in low-risk deciles and

overestimation in the high-risk decile. IBIS (E/O = 1.13, 95% CI = 1.05 to 1.23) showed a similar extent of overall miscalibration as iCARE-Lit and greater miscalibration in the high-risk deciles (Table 1, Figure 2; Supplementary Figure 4A, available online).

In PLCO, iCARE-Lit produced a similar overestimation of 5-year absolute risk as in the GS for women age 50 years or older. BCRAT and IBIS both underestimated absolute risk

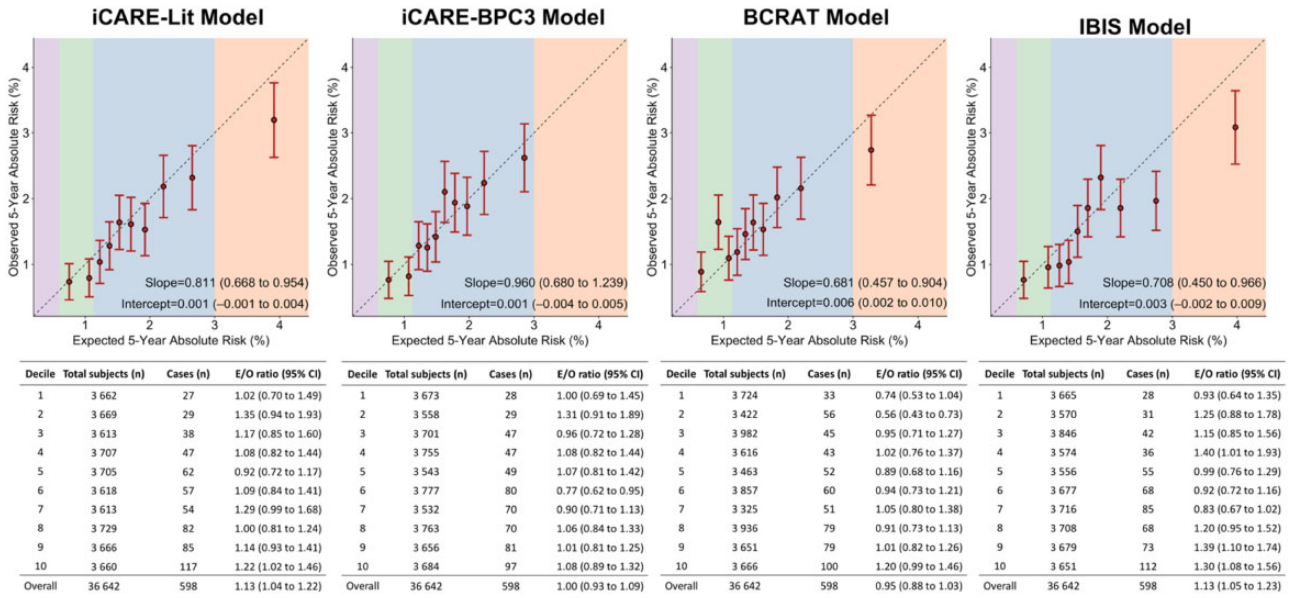


Figure 2. Absolute risk calibration of breast cancer risk prediction models in the GS cohort among women 50 years of age. The risk categories are based on absolute risk. The backgrounds of the plots are shaded to indicate the absolute risk threshold categories ($\leq 0.6\%$ in pink, $>0.6\%$ to $\leq 1.13\%$ in green, $>1.13\%$ to $\geq 3\%$ in blue, and $>3\%$ in orange). The 0.6% and 1.13% thresholds correspond to the average 5-year risk for US women age 40 years and 50 years, respectively. The 3% threshold is used by the US Preventive Services Task Force for recommending risk-lowering drugs, and 6% is used by the WISDOM trial as a threshold for very high risk. BCRAT = Breast Cancer Risk Assessment Tool; CI = confidence interval; E/O = expected-to-observed, GS = Generations Study; IBIS = International Breast Cancer Intervention Study; iCARE-BPC3 = Individualized Coherent Absolute Risk Estimation model based on Breast and Prostate Cancer Cohort Consortium; iCARE-Lit = iCARE model based on literature review.

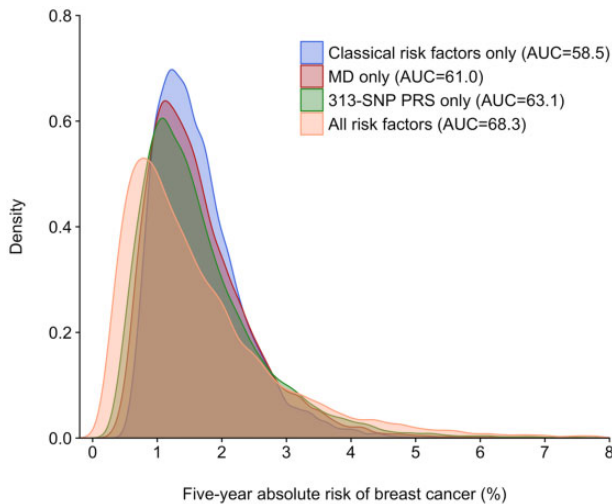


Figure 3. Five-year absolute risk projection for the general US population of white non-Hispanic women, ages 50–70 years. The classical risk factors correspond to the iCARE-BPC3 model. Classical risk factors include age at menarche, age at menopause, parity, age at first birth, height, alcohol intake, breast cancer family history, smoking status, body mass index, current HRT use, and ever HRT type. The projected AUCs reported are based on the relative risk score in that population and do not incorporate variation due to age. AUC = area under the curve; HRT = hormone replacement therapy; iCARE-BPC3 = Individualized Coherent Absolute Risk Estimation model based on Breast and Prostate Cancer Cohort Consortium; MD = mammographic breast density; PRS = polygenic risk score; SNP = single-nucleotide polymorphism.

(Supplementary Table 6 and Supplementary Figure 5A, available online). In both cohorts, discriminatory accuracy was lower when AUC was defined using the relative risk score, as opposed

to absolute risk. (Supplementary Figures 3B, 4B, and 5B, available online).

Breast Cancer Risk Projections

Figure 3 shows 5-year absolute risk projections in a target population of white non-Hispanic US women age 50–70 years (~30 million according to 2016 US Census). MD and 313-SNP PRS alone had higher AUCs compared with classical risk factors (based on iCARE-BPC3). An integrated model with classical risk factors, MD and PRS had the highest AUC of 68.3.

The classical risk factors could identify approximately 4.1 million women, representing 13.8% of the target population, at low risk ($<1.13\%$, corresponding to the average 5-year risk for 50-year-old US women) of invasive breast cancer, and 40 516 (8.2% of all cases) are expected to develop the disease within 5 years (Figure 4; Supplementary Table 7, available online). Integrating classical risk factors with MD and 313-SNP PRS is expected to increase the number of women to 12 million, and around 89 000 (17.7% of all cases) would be expected to develop the disease within 5 years. In the moderate- to high-risk group ($>3\%$ 5-year risk threshold based on US Preventive Services Task Force recommendation for risk-reducing therapies [41]), approximately 500 000 women, representing 1.7% of this population, could be identified based on classical risk factors, including approximately nearly 17 000 (3.4% of all cases) expected to develop the disease within 5 years (Figure 4; Supplementary Table 7, available online). Integrating with MD and 313-SNP PRS increases the number of women identified to 3.5 million, and among them, approximately 153 000 (~30% of all cases) would be expected to develop the disease within 5 years.

We projected that doubling the size of current breast cancer genome-wide association study (to around 300 000 patients and 300 000 controls) would yield additional discoveries and an

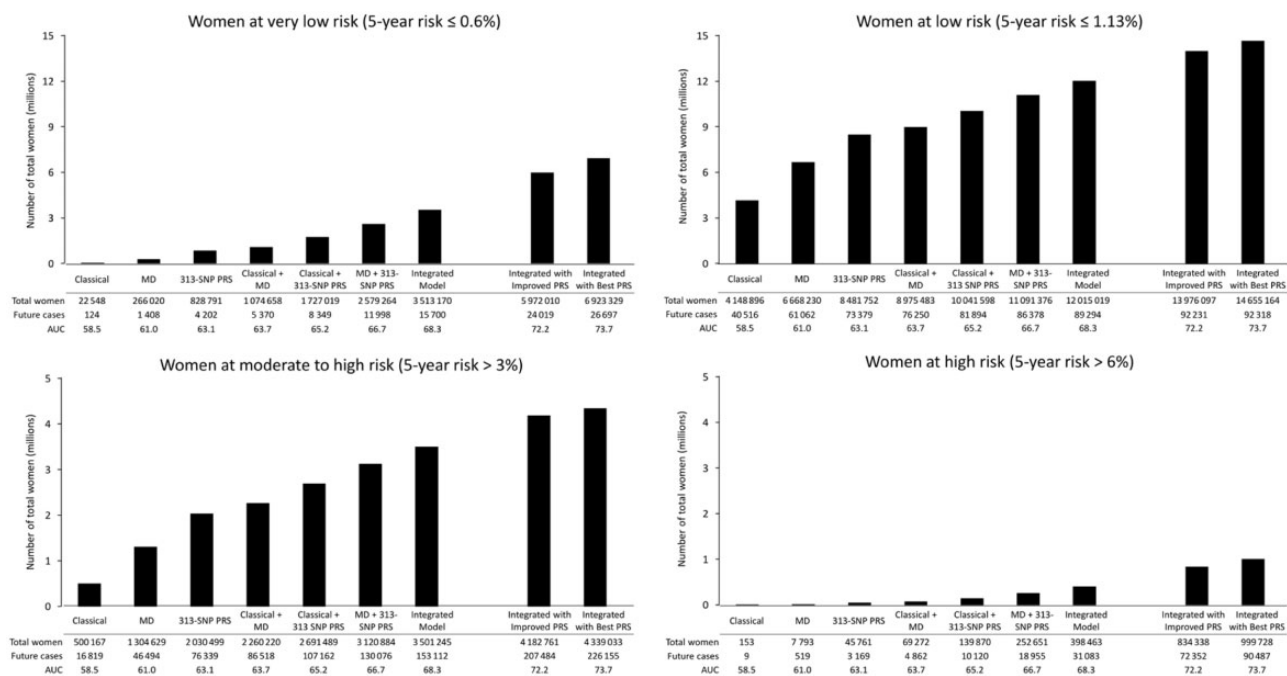


Figure 4. White non-Hispanic women age 50–70 years in the US population expected to be identified at different levels of risk of breast cancer according to four risk thresholds and the incident cases of invasive breast cancer who are expected to occur in these groups within a 5-year interval. The expected number of women is calculated using mid-2016 population estimates ($n = 30\,030\,821$) from the US Census Bureau, and the number of cases is calculated using the average predicted 5-year risk and the 2015 invasive breast cancer incidence rates from the Surveillance, Epidemiology, and End Results Program. The 0.6% and 1.13% thresholds correspond to the average 5-year risk for US women age 40 years and 50 years, respectively. The 3% threshold is used by the US Preventive Services Task Force for recommending risk-lowering drugs, and 6% is used by the WISDOM trial as a threshold for very high risk. Classical risk factors include age at menarche, age at menopause, parity, age at first birth, height, alcohol intake, breast cancer family history, smoking status, body mass index, current HRT use, and ever HRT type. The projected AUCs reported are based on the relative risk score in that population and do not incorporate variation due to age. AUC = area under the curve; HRT = hormone replacement therapy; MD = mammographic breast density; PRS = polygenic risk score; SNP = single-nucleotide polymorphism. iCARE-BPC3 = Individualized Coherent Absolute Risk Estimation model based on Breast and Prostate Cancer Cohort Consortium.

“improved” PRS with an AUC of 69.1 (Supplementary Table 7, available online). An integrated model with improved PRS could identify approximately 14 million women at low risk, and approximately 92 000 (~18% of all cases) would be expected to develop invasive breast cancer within 5 years (Figure 4; Supplementary Table 7, available online). In the moderate- to high-risk group, we could identify close to 4.2 million women, with approximately 207 000 (~40% of all cases) expected to develop the disease within 5 years. This is close to the risk stratification attained by the “best” theoretical PRS explaining 100% of the variability of polygenic risk from common variants (Figure 4; Supplementary Table 7, available online). Additionally, when incorporating the improved PRS, the relative increase in the numbers of women and cases identified increases as the risk threshold becomes more extreme at either end of the distribution (Figures 4; Supplementary Table 7, available online).

Evaluation of Net Benefit

We evaluated the theoretical net benefit for high-risk decisions based on the iCARE-BPC3 model and its extensions after the addition of PRS and MD in the US population of white non-Hispanic women age 50–70 years (Figure 5). At a 3% 5-year risk threshold used for recommendation for risk-lowering medications (41), there is virtually no net benefit for a model with classical risk factors alone, whereas the integrated models with the addition of PRS and MD show some net benefit for these women. However, none of the models show a net benefit for

risk-reducing interventions on women at the highest risk threshold (ie, above 6%).

Discussion

In this comparative analysis using data from a large population-based cohort, we showed that iCARE-based absolute risk models for invasive breast cancer with classical risk factors are similarly or better calibrated than previous models evaluated here, and that the addition of MD and PRS to classical risk factors can substantially improve risk stratification in the population.

Among women younger than 50 years, we found no substantial evidence of miscalibration of the relative risk for any of the models evaluated; however, we found some evidence for miscalibration of the 5-year absolute risk for all models except iCARE-Lit. This illustrates the challenges of validating models for absolute risk because, in addition to relative risk information, it requires information on population-based incidence rates and distribution of risk factors, ideally from the same time period as the validation study.

Among women age 50 years or older, we found no evidence for miscalibration of iCARE-BPC3 in terms of relative or absolute risk, whereas the other models overestimated absolute risk for women in the highest risk category. Although relative risk is reasonably well calibrated for both iCARE-based models, the better absolute risk calibration for iCARE-BPC3 compared with iCARE-Lit could be due to differences in specification of risk factors (ie, finer vs coarser categories of continuous risk factors).

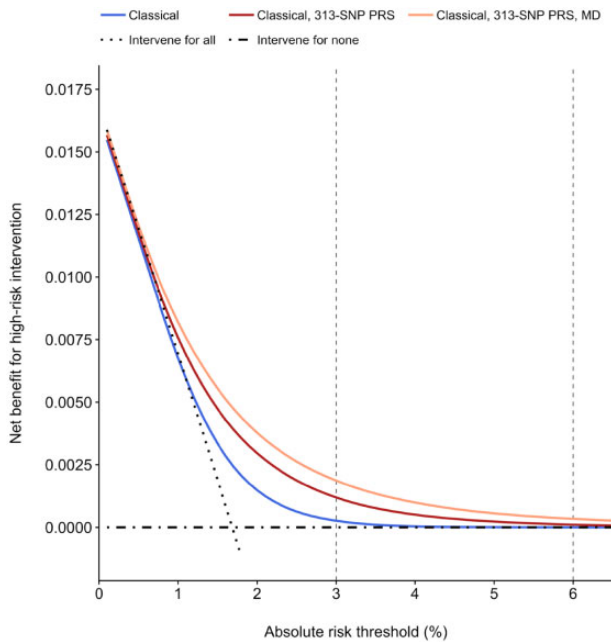


Figure 5. Projected net benefit of identifying white non-Hispanic women age 50–70 years with predicted absolute risk above a range of thresholds, with vertical lines representing 3% and 6% thresholds. The 3% threshold is used by the US Preventive Services Task Force for recommending risk-lowering drugs, and 6% is used by the WISDOM trial as a threshold for very high risk. The classical risk factors correspond to the iCARE-BPC3 model. Classical risk factors include age at menarche, age at menopause, parity, age at first birth, height, alcohol intake, breast cancer family history, smoking status, body mass index, current HRT use, and ever HRT type. Projections were made under the assumption of perfect calibration and log-normal distribution of risk in the population and using information on the spread (SD) of the risk score from the reference sample and distribution of age in the current US population. HRT = hormone replacement therapy; iCARE-BPC3 = Individualized Coherent Absolute Risk Estimation model based on Breast and Prostate Cancer Cohort Consortium; MD = mammographic breast density, PRS = polygenic risk score, SNP = single nucleotide polymorphism.

The IBIS and iCARE models include more information on classical risk factors than BCRAT (29,30), which has lower discriminatory accuracy. IBIS incorporates detailed family history information, considered important to identify high-risk women with extensive family history. However, comparisons of risk stratification across models in our study were limited by the miscalibration of BCRAT, IBIS, and iCARE-Lit for women age 50 years or older, particularly in high-risk deciles.

BCRAT has been extensively evaluated and is currently recommended for predicting breast cancer risk for US women undergoing mammographic screening (42,43). Whereas some studies found no evidence for miscalibration of BCRAT (44–47), others reported underestimation (48–50) or overestimation (47,51–53) of risk. Some studies reported improved calibration when using incidence and mortality rates from the same country and time period (49,50,53). Sensitivity analyses using rates closer to our validation population also indicated slight improvement in calibration (data not shown). Miscalibration may be due to model misspecification or not fully accounted-for differences in the risk factor distribution between the cohort and the underlying population.

Two validation studies in high-risk populations in the United States and United Kingdom found no evidence for miscalibration of IBIS (5,54,55). IBIS 10-year risk predictions have been found to be better calibrated in family-based studies

including average-to-high risk women, than the BCRAT predictions (5,6,56,57). IBIS and BCRAT both showed good absolute risk calibration in an Australian population of average-risk women (16). A recent prospective evaluation in a US-based integrated health-care system showed good calibration of IBIS 10-year absolute risks overall, but approximately 20% overestimation in the highest-risk decile (9). The current literature and our findings highlight the importance of external validation of risk models using multiple prospective studies to evaluate robustness of model performance across populations. Of note, validation studies often lack adequate size and rigorous methodology, making comparisons across studies difficult (58). There is a need for further robust validation in large studies to identify models adequate for clinical decision making (59).

Our study has several strengths. GS is a relatively recent population-based cohort including women with a wide age range. Moreover, the validation results of the iCARE-Lit model in PLCO further supported our overall conclusions. We evaluated model calibration overall and stratified by levels of risk. The latter is important because accurate classification of individuals at the extremes of risk is most relevant for risk-based prevention and screening. Second, we assessed model calibration by deciles of expected absolute risk and by the relative risk score. The former is commonly used in validation studies (5,44–46,48,53,54) because absolute risk is the relevant measure for clinical or public health applications. However, strong dependence of absolute risk on age makes the differences in model performance because of other risk factors less evident than comparisons using the relative risk score, which does not include age. We evaluated model calibration and discrimination with and without accounting for age. Most (5,45,48–51,54,55) but not all (44,46,53) previous validation studies of BCRAT and IBIS assessed model discrimination accounting for age. Such model discrimination statistics (eg, AUC) evaluated in a validation cohort may differ from those in the target population because of differences in risk factor distributions. Additionally, our results showed that small changes in overall measures (eg, AUC) derived from additional risk factors can result in substantial changes in the number of women at the extremes of risk distribution.

Limitations of our analyses include that not all risk factors were available in the validation cohorts. Additionally, further evaluation of iCARE models in cohorts more representative of populations in health-care systems where the models would be used is desirable. Moreover, we evaluated only short-term risk predictions, assuming risk factors remained constant over the prediction period. Validation of long-term risk will require further follow-up or additional studies, preferably accounting for time-varying risk factors and time-dependent associations. Our current model development and validation efforts focus on predicting risk of breast cancer for white non-Hispanic women. Our ongoing work will extend the models to nonwhite populations and include extensive classical risk factor information and an improved PRS (60,61).

For a given high-risk clinical decision, a well-calibrated model providing wider risk stratification is likely to have greater clinical utility. We have shown this using theoretical net benefit analyses in a target population; however, further assessment of the clinical utility of models will require identifying risk thresholds explicitly informed by benefits and costs of a specific intervention (20,62,63). Risk projections based on the integrated model assumed that classical risk factors, MD, and PRS act multiplicatively on disease risk. We accounted for known dependencies between classical risk factors, and previous studies support multiplicative effects of classical risk factors or MD

with PRS on disease risk (24,64,65). Risk projections based on models with MD accounted for its dependence on age and body mass index, but not on the other risk factors in the model that have weaker associations with MD (66). Thus, these should be considered only as projections, and such integrated models require independent prospective validation prior to consideration for clinical use. We derived PRS based on SNP odds ratios from genetic discovery studies (34,35). This may result in overestimation of risk stratification; however, based on previous assessment (67), this bias is likely small.

Updates to BCRAT and IBIS have added MD (7–11) and PRS (12–18) to the original models. However, only the addition of MD to IBIS has been prospectively validated and it showed some overestimation in the high-risk categories (9). Addition of an 18-SNP PRS to IBIS was shown to increase risk discrimination in a UK-based screening cohort, although accuracy of prediction was not prospectively evaluated (18). The Breast Cancer Surveillance Consortium risk model, which includes BI-RADS MD and a 76-SNP PRS, showed good calibration in a case-control study nested within a US-based cohort (68).

Further improvements in risk stratification could be achieved by incorporating additional risk factors and heterogeneity in risk factor associations by breast cancer subtypes. Ultimately, it is desirable to develop a comprehensive model, robustly validated in multiple populations with disparate risk factor information, applicable in populations with a wide range of underlying risk that provides reliable risk estimates based on subsets of risk factors depending on the clinical application (eg, risk assessment before or after mammography). The iCARE methodology facilitates this by providing a flexible risk modeling and validation tool with capabilities of handling missing risk factors.

In conclusion, we have demonstrated that iCARE models based on classical risk factors perform similarly to and, in some cases, better than established models for 5-year risk predictions of invasive breast cancer. Based on our projections, substantial improvements in risk stratification can be achieved with the addition of MD and PRS to classical risk factors. Such integrated risk models require further prospective empirical validation before broad clinical or research applications.

Funding

This work was supported by the Patient-Centered Outcomes Research Institute Award (ME-1602–34530), Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics (Z01CP010119), and the European Union's Horizon 2020 research and innovation program under grant agreement No. 633784 (B-CAST).

Notes

Affiliations of authors: Department of Biostatistics, Bloomberg School of Public Health (PPC, YZ, NC) and Department of Oncology, School of Medicine (NC), Johns Hopkins University, Baltimore, MD; Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda (ANW, TA, MHG, MG-C); Division of Genetics and Epidemiology (MNB, PC, MJS, MEJ, AJS), and Division of Breast Cancer Research (AJS), The Institute of Cancer Research, London, UK; Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, UK (NO).

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.

The authors have no conflicts of interest related to this work.

We would like to thank Adam R. Brentnall and Jack Cuzick for providing the IBIS Risk Evaluator software and for their comments to a draft of the manuscript. We would like to thank Celine Vachon for providing unpublished information on the association of mammographic breast density with invasive breast cancer and population distribution of mammographic density. This information was used in the risk projection calculations.

The Generations Study would like to thank Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study and the study participants, study staff, and the doctors, nurses, and other health-care staff and data providers who have contributed to the study. The Institute of Cancer Research acknowledges National Health Service funding to the National Institute for Health Research Biomedical Research Centre.

References

- 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(11):dju305.
- 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.
- Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast Cancer Res Treat*. 2012;132(2):365–377.
- Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinstian A. Risk prediction models of breast cancer: a systematic review of model performances. *Breast Cancer Res Treat*. 2012;133(1):1–10.
- Quante AS, Whittemore AS, Shriver T, Strauch K, Terry MB. Breast cancer risk assessment across the risk continuum: genetic and nongenetic risk factors contributing to differential model performance. *Breast Cancer Res*. 2012;14(6):R144.
- 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.
- Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat*. 2005;94(2):115–122.
- Chen J, Pee D, Ayyagari R, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J Natl Cancer Inst*. 2006;98(17):1215–1226.
- Brentnall AR, Cuzick J, Buist DSM, Bowles E. Long-term accuracy of breast cancer risk assessment combining classic risk factors and breast density. *JAMA Oncol*. 2018;4(9):e180174.
- 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.
- Warwick J, Birke H, Stone J, et al. Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: findings from the placebo arm of the International Breast Cancer Intervention Study I. *Breast Cancer Res*. 2014;16(5):451.
- Allman R, Dite GS, Hopper JL, et al. SNPs and breast cancer risk prediction for African American and Hispanic women. *Breast Cancer Res Treat*. 2015;154(3):583–589.
- Cuzick J, Brentnall AR, Segal C, et al. Impact of a panel of 88 single nucleotide polymorphisms on the risk of breast cancer in high-risk women: results from two randomized Tamoxifen prevention trials. *J Clin Oncol*. 2017;35(7):743–750.
- Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst*. 2010;102(21):1618–1627.
- Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. *J Natl Cancer Inst*. 2009;101(13):959–963.
- Dite GS, MacInnis RJ, Bickerstaffe A, et al. Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women age under 50 years: Australian Breast Cancer Family Registry. *Cancer Epidemiol Biomarkers Prev*. 2016;25(2):359–365.
- Evans DG, Brentnall A, Byers H, et al. The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case-control study. *J Med Genet*. 2017;54(2):111–113.
- van Veen EM, Brentnall AR, Byers H, et al. Use of single-nucleotide polymorphisms and mammographic density plus classic risk factors for breast cancer risk prediction. *JAMA Oncol*. 2018;4(4):476–482.
- Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. 2018;19(9): doi: 10.1038/s41576-018-0018-x.

20. Pashayan N, Morris S, Gilbert FJ, Pharoah P. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: a life-table model. *JAMA Oncol.* 2018;4(11):1504–1510.
21. Shieh Y, Hu D, Ma L, et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res Treat.* 2016;159(3):513–525.
22. Chatterjee N, Chen YH, Maas P, Carroll RJ. Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *J Am Stat Assoc.* 2016;111(513):107–117.
23. Grill S, Ankerst DP, Gail MH, Chatterjee N, Pfeiffer RM. Comparison of approaches for incorporating new information into existing risk prediction models. *Stat Med.* 2017;36(7):1134–1156.
24. Kundu P, Tang R, Chatterjee N. Generalized meta-analysis for multiple regression models across studies with disparate covariate information. arXiv:1708.03818v2 [stat.ME]. 2018. <https://arxiv.org/abs/1708.03818v2>. Accessed June 10, 2019.
25. Maas P, Barrdahl M, Joshi AD, et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol.* 2016;2(10):1295–1302.
26. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet.* 2016;17(7):392–406.
27. Pal Choudhury P, Maas P, Wilcox A, et al. iCARE: an R package to build, validate and apply absolute risk models. bioRxiv. 2018. <https://doi.org/10.1101/079954>. Accessed June 10, 2019.
28. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2006;15(6):1159–1169.
29. 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.
30. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med.* 2004;23(7):1111–1130.
31. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128–138.
32. Rousson V, Zumbunn T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. *BMC Med Inform Decis Mak.* 2011;11(45):doi: 10.1186/1472-6947-11-45.
33. Zhang Z, Rousson V, Lee WC, et al. Decision curve analysis: a technical note. *Ann Transl Med.* 2018;6(15):308.
34. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92–94.
35. Milne RL, Kuchenbaecker KB, Michailidou K, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet.* 2017;49(12):1767–1778.
36. 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.
37. Zhang Y, Qi G, Park JH, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet.* 2018;50(9):1318–1326.
38. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet.* 2013;45(4):400–405, 405e1–e3.
39. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet.* 2002;31(1):33–36.
40. Shieh Y, Eklund M, Madlensky L, et al. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. *J Natl Cancer Inst.* 2017;109(5):dju290.
41. U.S. Preventive Services Task Force. Final recommendation statement: Breast cancer: medications for risk reduction. 2016. <https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/breast-cancer-medications-for-risk-reduction>. Accessed December 13, 2018.
42. American College of Obstetricians and Gynecologists. Practice Bulletin Number 179: Breast cancer risk assessment and screening in average-risk women. *Obstet Gynecol.* 2017;130(1):e1–e16.
43. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines) Breast Cancer Screening and Diagnosis. Version 1.2019. https://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf. Accessed June 11, 2019.
44. Decarli A, Calza S, Masala G, Specchia C, Palli D, Gail MH. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence-European Prospective Investigation into Cancer and Nutrition cohort. *J Natl Cancer Inst.* 2006;98(23):1686–1693.
45. Arrospe A, Forné C, Rué M, Torá N, Mar J, Baré M. An assessment of existing models for individualized breast cancer risk estimation in a screening program in Spain. *BMC Cancer.* 2013;13:587:doi: 10.1186/1471-2407-13-587.
46. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med.* 2008;148(5):337–347.
47. Nickson C, Procopio P, Velentzis LS, et al. Prospective validation of the NCI Breast Cancer Risk Assessment Tool (Gail Model) on 40,000 Australian women. *Breast Cancer Res.* 2018;20(1):155.
48. 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.
49. Schonfeld SJ, Pee D, Greenlee RT, et al. Effect of changing breast cancer incidence rates on the calibration of the Gail model. *J Clin Oncol.* 2010;28(14):2411–2417.
50. 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(5):358–366.
51. Banegas MP, Gail MH, LaCroix A, et al. Evaluating breast cancer risk projections for Hispanic women. *Breast Cancer Res Treat.* 2012;132(1):347–353.
52. Spiegelman D, Colditz GA, Hunter D, Hertzmark E. Validation of the Gail et al. model for predicting individual breast cancer risk. *J Natl Cancer Inst.* 1994;86(8):600–607.
53. Pastor-Barriuso R, Ascunce N, Ederra M, et al. Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. *Breast Cancer Res Treat.* 2013;138(1):249–259.
54. Powell M, Jamshidian F, Cheyne K, Nititham J, Prebil LA, Ereman R. Assessing breast cancer risk models in Marin County, a population with high rates of delayed childbirth. *Clin Breast Cancer.* 2014;14(3):212–220.e211.
55. 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.
56. Mitchell H Gail. Performance of BCRAT in high-risk patients with breast cancer. *Lancet Oncol.* 2019;20(6):e285.
57. Terry MB, Liao Y, Hopper JL, MacInnis RJ. Performance of BCRAT in high-risk patients with breast cancer – Authors' reply. *Lancet Oncol.* 2019;20(6):e286.
58. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol.* 2014;14(40):doi: 10.1186/1471-2288-14-40.
59. Garcia-Closas M, Chatterjee N. Assessment of breast cancer risk: which tools to use? *Lancet Oncol.* 2019;20(4):463–464.
60. NCI Division of Cancer Epidemiology and Genetics. Confluence Project. <https://dceg.cancer.gov/research/cancer-types/breast-cancer/confluence-project>. Accessed December 21, 2018.
61. National Cancer Institute. NCI Cohort Consortium. <https://epi.grants.cancer.gov/Consortia/cohort.html>. Accessed January 27, 2019.
62. Mandelblatt JS, Stout NK, Schechter CB, et al. Collaborative modeling of the benefits and harms associated with different U.S. breast cancer screening strategies. *Ann Intern Med.* 2016;164(4):215–225.
63. Trentham-Dietz A, Kerlikowske K, Stout NK, et al. Tailoring breast cancer screening intervals by breast density and risk for women age 50 years or older: collaborative modeling of screening outcomes. *Ann Intern Med.* 2016;165(10):700–712.
64. Rudolph A, Song M, Brook MN, et al. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. *Int J Epidemiol.* 2018;47(2):526–536.
65. Vachon CM, Scott CG, Tamimi RM, et al. Joint association of mammographic density adjusted for age and body mass index and polygenic risk score with breast cancer risk. *Breast Cancer Res.* 2019;21(1):68.
66. Alexeeff SE, Odo NU, McBride R, et al. Reproductive factors and mammographic density: associations among 24,840 women and comparison of studies using digitized film-screen mammography and full-field digital mammography. *Am J Epidemiol.* 2019;188(6):1144–1154.
67. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst.* 2015;107(5):dju036.
68. Vachon CM, Pankratz VS, Scott CG, et al. The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst.* 2015;107(5):dju397.