CANCER PREVENTION

original reports abstract

Integration of a Cross-Ancestry Polygenic Model With Clinical Risk Factors Improves Breast Cancer Risk Stratification

Placede T. Tshiaba, MS[1]; Dariusz K. Ratman, PhD[1]; Jiayi M. Sun, PhD[1]; Tate S. Tunstall, PhD[1]; Brynn Levy, PhD[1,2]; Premal S. Shah, PhD[1]; Jeffrey N. Weitzel, MD[3]; Matthew Rabinowitz, PhD[1,3]; Akash Kumar, MD, PhD[1]; and Kate M. Im, PhD[1]

PURPOSE To develop and validate a cross-ancestry integrated risk score (caIRS) that combines a cross-ancestry polygenic risk score (caPRS) with a clinical estimator for breast cancer (BC) risk. We hypothesized that the caIRS is a better predictor of BC risk than clinical risk factors across diverse ancestry groups.

METHODS We used diverse retrospective cohort data with longitudinal follow-up to develop a caPRS and integrate it with the Tyrer-Cuzick (T-C) clinical model. We tested the association between the caIRS and BC risk in two validation cohorts including > 130,000 women. We compared model discrimination for 5-year and remaining lifetime BC risk between the caIRS and T-C and assessed how the caIRS would affect screening in the clinic.

RESULTS The caIRS outperformed T-C alone for all populations tested in both validation cohorts and contributed significantly to risk prediction beyond T-C. The area under the receiver operating characteristic curve improved from 0.57 to 0.65, and the odds ratio per standard deviation increased from 1.35 (95% CI, 1.27 to 1.43) to 1.79 (95% CI, 1.70 to 1.88) in validation cohort 1 with similar improvements observed in validation cohort 2. We observed the largest gain in positive predictive value using the caIRS in Black/African American women across both validation cohorts, with an approximately two-fold increase and an equivalent negative predictive value as the T-C. In a multivariate, age-adjusted logistic regression model including both caIRS and T-C, caIRS remained significant, indicating that caIRS provides information over T-C alone.

CONCLUSION Adding a caPRS to the T-C model improves BC risk stratification for women of multiple ancestries, which could have implications for screening recommendations and prevention.

JCO Precis Oncol 7:e2200447. © 2023 by American Society of Clinical Oncology

INTRODUCTION

Accurate individualized breast cancer (BC) risk estimates are critical for identifying women eligible for preventive and early detection interventions. BC risk estimates are based on combinations of clinical and epidemiologic risk factors, such as age, family history (FH) of cancer, age at menarche, and body mass index.[1,2] Some models also account for genetic risk factors, typically pathogenic variants (PVs) in high-risk (including *BRCA1*, *BRCA2*, and *PALB2*) and moderate-risk genes (such as *ATM* and *CHEK2*).[3-8] However, these variants are rare in the general population and only account for approximately 6% of all BCs and 2.6% when considering only *BRCA1* and *BRCA2*.[5,9] To more adequately evaluate the contribution of genetic factors to BC risk, polygenic risk scores (PRSs) have been developed on the basis of common variants associated with BC in genome-wide association studies (GWASs). Individually, each of these variants confers only a small increase in risk, but collectively, as a PRS, they account for considerable BC susceptibility.[10-15]

Recent studies have demonstrated that use of PRSs enables more effective risk stratification and improves the risk-benefit ratio in population-wide BC screening.[16,17] Wolfson et al[17] showed that PRS information more effectively stratified women for BC risk than BC FH or PV alone. PRSs can be integrated with traditional clinical risk models, which include cancer FH, to potentially improve BC risk assessment. To date, studies have consistently demonstrated improved BC risk stratification with integrated models versus clinical models alone.[2,18-25]

However, most polygenic models have been developed and validated in women of European ancestry.[2,12,18,21,22,26-34] When they have been generalized to non-European populations, these PRSs typically do not provide comparable risk stratification.[34,35] These findings highlight the need

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on December 22, 2022 and published at ascopubs.org/journal/po on February 21, 2023: DOI https://doi.org/10.1200/PO.22.00447

CONTEXT

Key Objective

To develop a risk prediction model using polygenic risk scores that improves breast cancer (BC) risk stratification over a standard-of-care tool, the Tyrer-Cuzick (T-C) clinical model, that can be applied to women of diverse ancestries.

Knowledge Generated

The cross-ancestry integrated risk score is significantly associated with BC risk and is well-calibrated in a validation set consisting of women of European and non-European ancestries. Relative to the T-C model alone, the cross-ancestry integrated risk score improved risk stratification for women of diverse ancestries.

Relevance

Addition of a cross-ancestry polygenic risk score to the T-C clinical model provides more accurate, individualized BC risk estimates for unaffected women of multiple ancestries for screening and prevention.

to develop and validate PRSs on the basis of variants identified in women of diverse ancestries.[20,23,24,35,36]

Recently, we described a method to generate a cross-ancestry PRS (caPRS) that improved risk stratification among women of five ancestries.[37] Here, we improved on our previous caPRS and integrated the caPRS with the Tyrer-Cuzick (T-C) clinical risk model to generate a cross-ancestry integrated risk score (caIRS). We hypothesized that the caIRS would be a better predictor of BC than the T-C score alone.

METHODS

Study Populations

We used genotype and phenotype data from multiple cohorts to develop and validate the caIRS. These cohorts included the UK Biobank (UKB),[38] the Women's Health Initiative (WHI, dbGAP study phs000200.v12.p3),[39] the GWAS of Breast Cancer in the African Diaspora (ROOT, dbGAP study phs000383.v1.p1),[40] and the Multiethnic Cohort (dbGAP study phs000517.v3.p1).[41] See the Data Supplement for more details on the inclusion/exclusion criteria.

We divided the cohorts into four study cohorts for development and validation of the caIRS (Data Supplement). Development cohorts 1 and 2 comprised 10,927 women with BC (cases) and 114,390 women unaffected by BC (controls) from the UKB (a subset), Multiethnic Cohort, and ROOT (Data Supplement). Validation cohort 1 comprised 25,284 women from the WHI, and validation cohort 2 comprised the remaining 119,187 women in the UKB who were not included in the development cohorts (Table 1). BC cases in the validation cohorts were diagnosed after initial assessment. Women diagnosed with BC within 5 years of initial assessment in the prospective validation cohorts were used to assess 5-year risk performance.

Genotype Imputation

We prephased genotype data from microarrays in each study using SHAPEIT2 and imputed unobserved genotypes using IMPUTE2 using either the 1000 Genomes Project phase III reference or UK10K reference[42,43] (further details are provided in the Data Supplement). For sites that could not be imputed in this way, we obtained a population-specific allele frequency from gnomAD v3.1.1 to estimate the average contribution of the variant.

Clinical Risk

We obtained cancer FH and risk factor information for participants in each study and then calculated absolute risk (5-year and remaining lifetime risk) estimates using T-C version 8.[44] We coded missing data according to the specifications of the model.

Statistical Methods

Cross-ancestry PRS. We evaluated the performance of multiple BC polygenic models across individuals in each of five ancestry groups (European, African, South Asian, East Asian, and Admixed American) using multivariate logistic regression models adjusted for age at assessment/diagnosis, first-degree FH of BC, and personal history of ovarian cancer (when available). To examine the performance of each model, we estimated the odds ratio (OR) per standard deviation (SD) and area under the receiver operating characteristic curve (AUC) in development cohort 1 and included the best performing model for each ancestry in the caPRS (Data Supplement).

We defined the caPRS as a linear combination of the best performing PRS model for each ancestry group weighted by the effect size and fractional ancestry. We evaluated the impact of FH of BC in first-degree relatives on caPRS by fitting an additional interaction term to the models. Further details can be found in the Data Supplement.

Associations between caPRS and T-C variables. We assessed the association between the caPRS and T-C variables in development cohort 2 using a linear regression model in which we predicted caPRS as a function of each of the T-C variables. We examined the regression coefficients and *P* values associated with the *F* statistic in these models.

TABLE 1. Validation Cohort Characteristics

Self-Reported Ethnicity	Total, No. (%)	5-Year Cases, No. (%)	5-Year Controls, No. (%)	Lifetime Cases, No. (%)	Lifetime Controls, No. (%)
Validation cohort 1	25,284 (100.0)	538 (2.1)	24,746 (97.8)	1,997 (7.9)	23,287 (92.1)
African American/Black	7,650 (30.3)	146 (27.1)	7,504 (30.3)	612 (30.6)	7,038 (30.2)
White	14,426 (57.1)	334 (62.1)	14,092 (56.9)	1,192 (59.7)	13,234 (56.8)
Hispanic/Latino	3,208 (12.7)	58 (10.8)	3,150 (12.7)	193 (9.7)	3,015 (12.9)
Age range, years	49-81	50-79	49-81	50-79	49-81
Median age, years	64	65	64	63	65
≥ 1 FDR with BC	3,748 (14.8)	98 (18.2)	3,650 (14.7)	387 (19.4)	3,361 (14.4)
caPRS range	−1.7 to 1.7	−1.1 to 1.7	−1.7 to 1.7	−1.3 to 1.7	−1.7 to 1.7
caPRS median	0.41	0.39	0.41	0.40	0.41
Validation cohort 2	119,187 (100.0)	3,628 (3.0)	115,559 (97.0)	8,115 (6.8)	111,072 (93.2)
Black	1,239 (1.0)	19 (0.5)	1,220 (1.1)	56 (0.7)	1,183 (1.1)
Chinese	468 (0.4)	6 (0.2)	462 (0.4)	16 (0.2)	452 (0.4)
White/White British	104,661 (87.8)	3,242 (89.4)	101,419 (87.8)	7,255 (89.4)	97,406 (87.7)
(South)Asian or Asian British	1,327 (1.1)	29 (0.8)	1,298 (1.1)	75 (0.9)	1,252 (1.1)
Others	11,492 (9.6)	332 (9.2)	11,160 (9.7)	713 (8.8)	10,779 (9.7)
Age range, years	40-70	40-70	40-70	40-70	40-70
Median age, years	57	59	57	59	57
≥ 1 FDR with BC	13,095 (11.0)	588 (16.2)	12,507 (10.8)	1,290 (15.9)	11,805 (10.6)
caPRS range	−4.3 to 4.6	−3.2 to 4.6	−4.3 to 4.4	−3.5 to 4.6	−4.3 to 3.8
caPRS median	−0.12	0.37	−0.13	0.33	−0.15

NOTE. The validation cohorts consisted of women from the Women's Health Initiative (validation cohort 1) and the UK Biobank (validation cohort 2).

Abbreviations: BC, breast cancer; caPRS, cross-ancestry polygenic risk score; FDR, first-degree relative.

Cross-ancestry integrated risk score. We estimated the effect size associated with the caPRS in development cohort 2 using a multivariate logistic regression that included caPRS, age at assessment/diagnosis, FH of BC, personal and FH of ovarian cancer (when available), and cohort. We separated data from the women unaffected by BC into 12 groups on the basis of age and first-degree FH of BC and calculated an adjustment constant for each group. We then calculated the 5-year and remaining lifetime risks on the basis of the T-C clinical model and caIRS (T-C combined with the caPRS). Further details are provided in the Data Supplement.

Validation studies. We evaluated the performance of the caIRS in two independent validation cohorts of women who were not used for model development. Associations with BC risk were evaluated in terms of *P* values and OR per SD from multivariate logistic regression models adjusted for age. The AUC was used to assess model discrimination. All analyses were performed using R Statistical Software (v4.1.0 or higher). Two-sided *P* values were calculated from likelihood ratio chi-square test statistics and reported at a significance level of .05. All analyses were performed for 5-year and remaining lifetime BC risk.

Reclassification of risk. To assess how the caIRS would affect screening in the clinic, we examined sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for 5-year risk using ≥ 3% as the high-risk threshold on the basis of the US Preventive Services Task Force recommendation for risk-lowering medications[45] and remaining lifetime risk using ≥ 20% as the high-risk threshold. No adjustments were made for competing mortality in our analysis.

RESULTS

Cross-Ancestry PRS

We built multiple polygenic models by training on multiple external GWAS sources (Data Supplement) and compared their performances against several published models. To ensure uniformity of evaluation, the performance of each PRS was evaluated on the same group of testing individuals (see the Data Supplement for cohort information). For each ancestral group, we selected the model that showed the best performance (Data Supplement). The AUC ranged from 0.636 to 0.684 for women of European descent and from 0.568 to 0.591 for women of African descent (Data Supplement). The scores for each ancestral model were

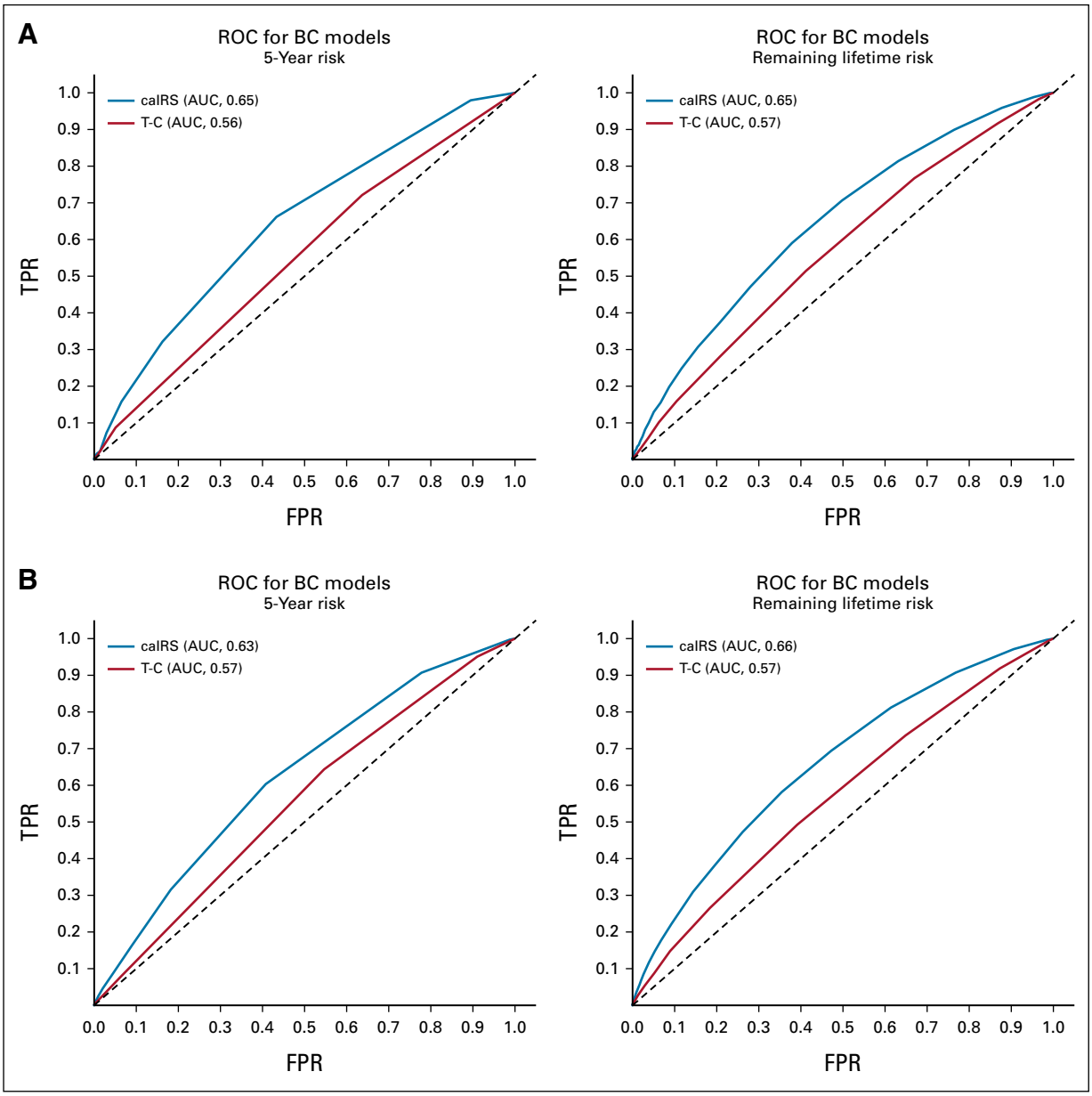

FIG 1. Performance of calRS compared with T-C alone in predicting 5-year and remaining lifetime risk of BC for women in the validation cohorts. AUC, area under the receiver operating characteristic curve; BC, breast cancer; calRS, cross-ancestry integrated risk score; FPR, false-positive rate; ROC, receiver operating characteristic curve; T-C, Tyrer-Cuzick; TPR, true-positive rate.

combined and weighted by fraction ancestry and effect size into a single caPRS. We evaluated the performance of the caPRS in development cohort 2. The caPRS was significantly associated with BC ($P = 2.3 \times 10^{-143}$), and the caPRS quantile was correlated with odds of BC (Data Supplement).

In addition, we examined the performance of the caPRS in women with and without first-degree FH of BC. We found no significant interaction between the caPRS and first-degree FH of BC for all population groups (P values ranging from 0.068 to 0.91), indicating that the association between the caPRS and BC is independent of FH of BC and that the caPRS is associated with BC risk in women with and without FH of BC. This is consistent with a recent publication that examined FH and PRS across > 20 diseases.[46]

Associations Between caPRS and T-C Variables

We examined associations between the caPRS and T-C model clinical factors in development cohort 2 using a linear regression model. The caPRS was significantly associated with age ($P = 2.6 \times 10^{-38}$) and first-degree FH of BC ($P = 4 \times 10^{-49}$). After Bonferroni correction for multiple testing, the caPRS was not correlated with any other T-C model clinical factors (Data Supplement).

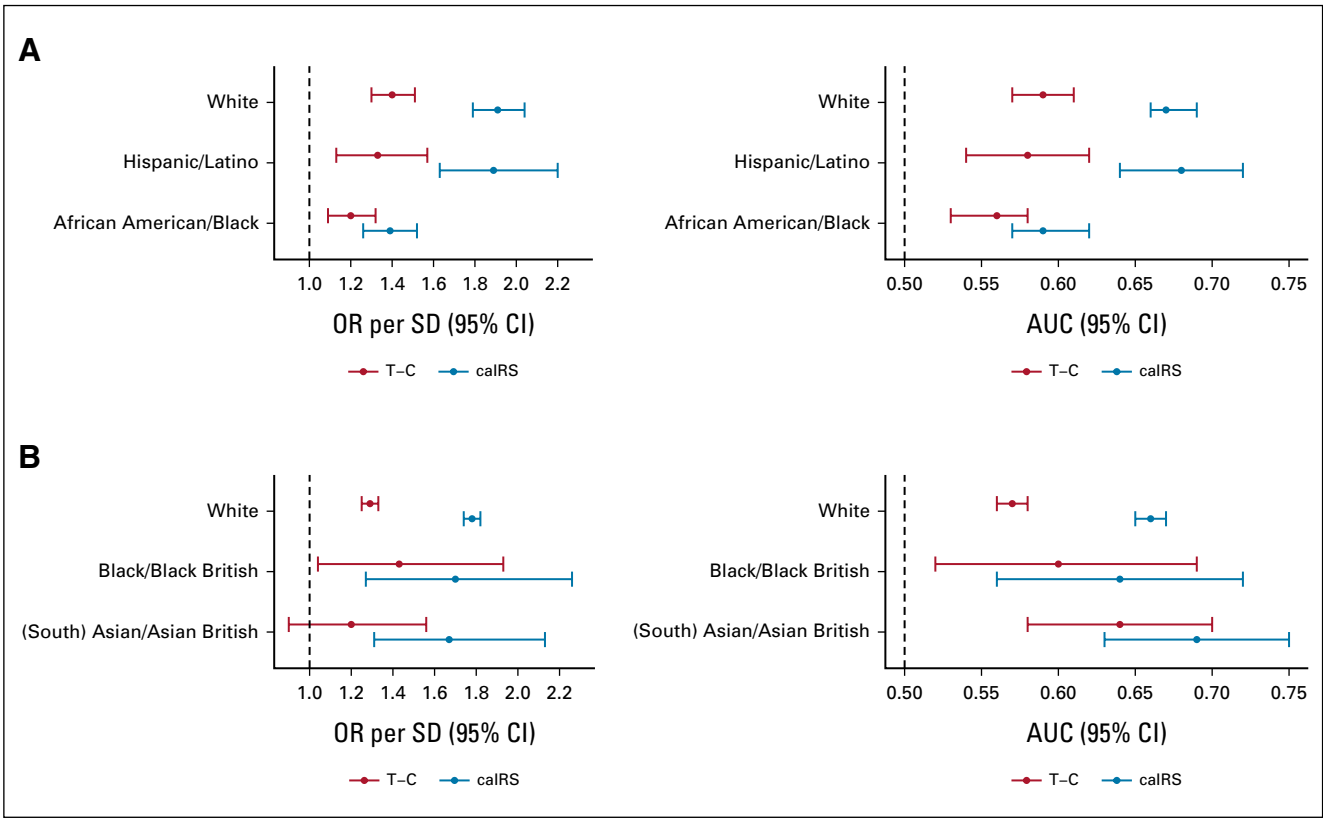

FIG 2. Performance of calRS compared with the T-C clinical model alone in predicting remaining lifetime risk of breast cancer across self-reported ethnicities in validation cohort 1 (A) and validation cohort 2 (B). Data for Chinese women in validation cohort 2 are not shown because of small sample size, but an improvement is observed (Data Supplement). AUC, area under the receiver operating characteristic curve; calRS, cross-ancestry integrated risk score; OR, odds ratio; SD, standard deviation; T-C, Tyrer-Cuzick.

Validation Studies

Validation cohort 1 comprised 25,284 women from the WHI, including 1,997 (7.9%) women diagnosed with BC after initial assessment, with a mean follow-up time of 17 years. Validation cohort 2 comprised the remaining 119,187 women in the UKB, including 8,115 (6.8%) diagnosed with BC, with a mean follow-up time of 6 years. Table 1 shows the subgroup sample sizes, after exclusions, in each validation cohort. The caPRS was significantly associated with 5-year and remaining lifetime risk in both validation cohorts (Data Supplement).

Figures 1 and 2 summarize relative performance improvements with the integrated risk score. The calRS was significantly associated with 5-year and remaining lifetime risk of BC (Data Supplement). In both validation cohorts, the calRS yielded significant improvements over the conventional T-C with an AUC of 0.65 (95% CI, 0.64 to 0.66) and an OR per SD of 1.79 (95% CI, 1.70 to 1.88) in validation cohort 1 and an AUC of 0.66 (95% CI, 0.65 to 0.66) and an OR per SD of 1.75 (95% CI, 1.71 to 1.79) in validation cohort 2 for remaining lifetime risk (Table 2 and Data Supplement). In addition, when both the T-C and calRS were included in a multivariate, age-adjusted logistic regression model, calRS remained significant, indicating that the calRS provides information over the T-C alone ($P = 2.1 \times 10^{-92}$ and $< 10^{-324}$ for validation cohorts 1 and 2, respectively).

The calRS outperformed the T-C alone for all subgroups tested in both validation cohorts, with the largest improvement in remaining lifetime risk observed in Hispanic women in validation cohort 1 where the OR per SD increased from 1.31 (95% CI, 1.10 to 1.54) to 1.88 (95% CI, 1.62 to 2.19) and the AUC increased from 0.58 to 0.68 and in European women in validation cohort 2 where the OR per SD increased from 1.29 (95% CI, 1.25 to 1.33) to 1.78 (95% CI, 1.74 to 1.82) and the AUC increased from 0.57 to 0.66 (Fig 2 and Data Supplement).

Overall and for all self-reported ancestry groups, the calRS was well calibrated across all deciles (Data Supplement). Overall, in validation cohort 1, the *P* values associated with the Hosmer-Lemeshow test statistic were 0.16 and 0.44 for 5-year and remaining lifetime BC risk, respectively. In addition, the expected-to-observed ratio at the highest risk decile were 1.07 (95% CI, 0.89 to 1.29) and 1.06 (95% CI, 0.96 to 1.17) for calRS for 5-year and remaining lifetime BC risk, respectively.

Reclassification of Risk

Across all women and for women in each population group, we saw an increase in sensitivity and a slight decrease in

TABLE 2. Performance of T-C Alone and caIRS in Predicting 5-Year and Remaining Lifetime Risk of Breast Cancer Across All Women in the Validation Cohorts

Model	5-Year Risk			Remaining Lifetime Risk		
	OR Per SD (95% CI)	*P*	AUC (95% CI)	OR per SD (95% CI)	*P*	AUC (95% CI)
Validation cohort 1						
T-C	1.18 (1.08 to 1.28)	1.9×10^{-4}	0.56 (0.54 to 0.59)	1.35 (1.27 to 1.43)	2.8×10^{-24}	0.57 (0.56 to 0.59)
caIRS	1.62 (1.50 to 1.76)	1.8×10^{-30}	0.65 (0.62 to 0.67)	1.79 (1.70 to 1.88)	4.5×10^{-110}	0.65 (0.64 to 0.66)
Validation cohort 2						
T-C	1.11 (1.08 to 1.15)	2.3×10^{-10}	0.57 (0.56 to 0.58)	1.28 (1.25 to 1.32)	5.8×10^{-75}	0.57 (0.57 to 058)
caIRS	1.46 (1.42 to 1.51)	9.8×10^{-128}	0.63 (0.62 to 0.64)	1.75 (1.71 to 1.79)	$< 10^{-324}$	0.66 (0.65 to 0.66)

Abbreviations: AUC, area under the receiver operating characteristic curve; caIRS, cross-ancestry integrated risk score; OR, odds ratio; SD, standard deviation; T-C, Tyrer-Cuzick.

specificity for 5-year risk estimation. However, PPV and, to a lesser extent, NPV increased in all groups compared with the T-C (Data Supplement). The largest gain in PPV using the caIRS occurred in Hispanic women, where we observed a nearly three-fold increase relative to that of the T-C while having a similar NPV. Overall, 18.3% and 30.6% of women who developed BC within 5 years of initial assessment were predicted to be at high risk (≥ 3%) by T-C and caIRS, respectively. Across the two methods, 8.8% of women had discordant classifications of risk.

We observed a similar trend for remaining lifetime risk for women in validation cohort 1 with a couple of exceptions. Despite an increase in AUC and sensitivity of the caIRS in Hispanic women, we observed a small drop in PPV across all age groups (Data Supplement). We observed the largest gain in PPV using the caIRS in African/African American women, an approximately two-fold increase, with an NPV equivalent to that of the T-C. The caIRS was able to correctly identify 6.2% of cases as high risk across all women age younger than 70 years in validation cohort 1, a 3-fold increase over the T-C alone. Both measures were able to identify more affected younger women (50-59 years) as being at high risk.

Overall, 8.8% and 6.7% of women across both validation cohorts were classified differently on the basis of 5-year and remaining lifetime risk, respectively, using the caIRS compared with the T-C alone (Figure 3 and Data Supplement).

DISCUSSION

Addition of single nucleotide polymorphism (SNP)-based PRSs to existing clinical BC risk prediction models has been shown to enhance discriminatory power.[2,18,19,21,22,47-49] Most of these PRSs were developed in women of European ancestry; by contrast, our caPRS is informative for women of multiple ancestries.

This study built a framework to efficiently iterate over polygenic models and addressed the limitations of some previous studies. First, our approach does not rely on self-reported ancestry (which can be inaccurate) in normalizing risk scores or a one-size-fits-all polygenic model. Second, our approach is not limited to a small set of SNPs, allowing us to add and remove variants as new data become available. Finally, both our development and validation cohorts were chosen to include minorities so that we could optimize risk stratification across women of diverse ancestries in parallel. Although our caIRS is significantly associated with risk of BC and improves on the performance of the T-C model in women of diverse ancestries, we see the best performance in women of European ancestry.

When developing a new method of assessing disease risk, maximizing sensitivity to identify women at higher risk of developing BC can lead to a loss of specificity. This tradeoff can be attenuated with proper calibration such that combining the PRS with the clinical model does not severely overestimate risk. In this study, we observed a small drop in specificity (1%-4% depending on the population) compared with the T-C when examining women of all ages. The small drop in PPV and NPV for remaining lifetime risk in Hispanic women despite an increase in AUC and OR could indicate that a 20% threshold is not optimal for the integrated risk in this population.

Here, we demonstrated that the caIRS, which combines a caPRS derived from five ancestry-specific polygenic models with the T-C risk model, significantly improved BC risk stratification across women of all ancestries tested relative to the T-C risk model alone. Overall, we found that approximately 8%-9% and 2%-6% of women would be reclassified as having a 3% or greater 5-year risk and 20% or greater remaining lifetime risk of BC, respectively, according to caIRS. If guidelines included the use of PRSs, these women would now qualify for consideration of risk-lowering medication and/or enhanced surveillance according to US Preventive Services Task Force and American Cancer Society recommendations.[45,50] In addition, approximately 0.5%-1.5% would be reclassified as having < 20% remaining lifetime risk of BC.

This study has some limitations because of the cohorts used for development and validation. In all cohorts, the

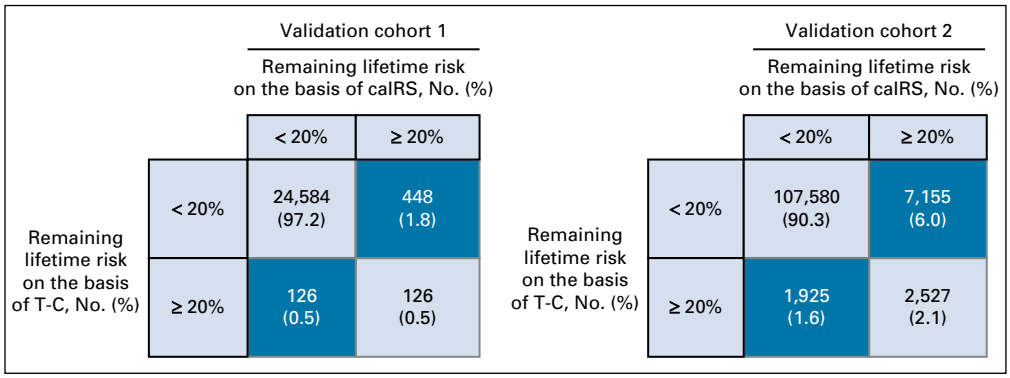

FIG 3. Distribution of women with high (≥ 20%) and average/low (< 20%) remaining lifetime risk in validation cohort 1 and validation cohort 2 on the basis of the calRS and T-C alone. See the Data Supplement for the distribution in individual populations. calRS, cross-ancestry integrated risk score; T-C, Tyrer-Cuzick.

representation of Asian women was much lower than other ancestral groups. We hope to address this by analyzing larger data sets from women of East and South Asian ancestries. Single gene variant information was not available for all cohorts. We do not expect this to significantly affect the results because the majority of women will not have PVs. We included White participants in the WHI in validation cohort 1; however, it should be noted that a subset of these women were part of a much larger GWAS study that was used in the development of models included in the caPRS. It may be expected that the performance of the calRS would be elevated in this group; however, the performance of the self-reported White women in validation cohort 2 is similar to that of self-reported White women in validation cohort 1.

The clinical information available for women in validation cohort 2, particularly for FH, was less detailed than that expected for the T-C model. Thus, the improvement in the accuracy and predictive ability of the calRS compared with that of the T-C may be somewhat inflated. However, the overall performance (AUC) of the calRS remained consistent in both cohorts, suggesting that the calRS is applicable even when detailed FH information is unavailable. In addition, the major change in v8 of the T-C model was the addition of breast density. We did not have that information for the women in our cohorts and therefore relied on default values in the model. This can be addressed by testing the calRS in a cohort of women with extremely detailed clinical records.

Finally, we validated our calRS in population cohorts, not women receiving hereditary cancer screening. We found that the caPRS works equally well in women with (n = 4,081; OR per SD, 1.74 [95% CI, 1.57 to 1.93]; $P = 5.5 \times 10^{-27}$) and without FH of BC (n = 32,678; OR per SD, 1.73 [95% CI, 1.65 to 1.82]; $P = 9.1 \times 10^{-119}$) in development cohort 2. This suggests that the calRS will perform as well in populations with higher percentages of FH. In addition, the women in validation cohort 1 are older, postmenopausal women, which likely does not represent the group of women for whom both measures are most effective at identifying those at high risk and who may benefit most from early/additional surveillance. To address both these limitations, we plan to test the calRS in younger, unaffected women seeking hereditary cancer screening.

Implementation of a more comprehensive risk model, as discussed here, may enable personalized screening and risk reduction strategies for women who are at high risk of BC, regardless of ancestry.

AFFILIATIONS

[1]MyOme Inc, Menlo Park, CA
[2]Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY
[3]Natera Inc, San Carlos, CA

CORRESPONDING AUTHOR

Kate M. Im, PhD, 930 Brittan Avenue, San Carlos, CA 94070; e-mail: kate@myome.com.

DISCLAIMER

This manuscript was not prepared in collaboration with investigators of the GWAS of Breast Cancer in the African Diaspora and does not necessarily reflect the opinions or views of University of Chicago, or NCI. This manuscript was not prepared in collaboration with investigators of the WHI, has not been reviewed and/or approved by the Women's Health Initiative (WHI), and does not necessarily reflect the opinions of the WHI investigators or the NHLBI.

PRIOR PRESENTATION

A version of the caPRS was presented as a poster at the 2022 ASCO Annual Meeting, Chicago, IL, June 3-7, 2022. The calRS was presented as a poster at the ASHG 2022 Annual Meeting, Los Angeles, CA, October 25-29, 2022.

SUPPORT

Supported by MyOme Inc. Funding for UK10K was provided by the Wellcome Trust under Award No. WT091310. Medical writing and editorial support was provided by Lauren Dembeck, PhD, funded by MyOme. The Multiethnic Cohort and the genotyping in this study were funded by grants from the National Institute of Health (CA63464, CA54281, CA098758, CA132839, and HG005922) and the Department of Defense Breast Cancer Research Program (W81XWH-08-1-0383). The WHI program was funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, and US Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. WHI PAGE was funded through the NHGRI Population Architecture Using Genomics and Epidemiology (PAGE) network (Grant No. U01 HG004790). Funding support for WHI GARNET was provided through the NHGRI Genomics and Randomized Trials Network (GARNET; Grant No. U01 HG005152). Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH Genes, Environment and Health Initiative (GEI; U01 HG004424). The Women's Health Initiative Sequencing Project (WHISP) was funded by Grant No. RC2 HL102924. This study was part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI Grant Nos. RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO), and RC2 HL102924 (WHISP). Exome sequencing was performed through NHLBI grant Nos. RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). Funding for WHI SHARe genotyping was provided by NHLBI Contract N02-HL-64278.

DATA SHARING STATEMENT

The data sets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accessions phs000383.v1.p1, phs000200.v12.p3, and phs000517.v3.p1.

AUTHOR CONTRIBUTIONS

Conception and design: Placede T. Tshiaba, Premal S. Shah, Matthew Rabinowitz, Akash Kumar, Kate M. Im
Financial support: Matthew Rabinowitz
Administrative support: Matthew Rabinowitz
Provision of study materials or patients: Matthew Rabinowitz
Collection and assembly of data: Placede T. Tshiaba, Dariusz K. Ratman, Matthew Rabinowitz, Kate M. Im
Data analysis and interpretation: Placede T. Tshiaba, Dariusz K. Ratman, Jiayi M. Sun, Tate S. Tunstall, Brynn Levy, Jeffrey N. Weitzel, Matthew Rabinowitz, Akash Kumar, Kate M. Im
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Placede T. Tshiaba
Employment: MyOme
Stock and Other Ownership Interests: MyOme
Research Funding: MyOme
Patents, Royalties, Other Intellectual Property: Several patents in the course of work at MyOme
Travel, Accommodations, Expenses: MyOme

Dariusz K. Ratman
Employment: MyOme
Stock and Other Ownership Interests: Roche, MyOme
Consulting or Advisory Role: MyOme
Research Funding: MyOme

Jiayi M. Sun
Employment: MyOme
Stock and Other Ownership Interests: MyOme

Tate S. Tunstall
Employment: MyOme
Stock and Other Ownership Interests: MyOme

Brynn Levy
Stock and Other Ownership Interests: Natera, MyOme
Consulting or Advisory Role: Igenomix, Natera, MyOme, Thermo Fisher Scientific Biomarkers
Expert Testimony: University of San Francisco, Illumina

Premal S. Shah
Employment: MyOme
Leadership: MyOme
Stock and Other Ownership Interests: MyOme
Consulting or Advisory Role: InterVenn Biosciences
Travel, Accommodations, Expenses: MyOme

Jeffrey N. Weitzel
Employment: Natera
Stock and Other Ownership Interests: Natera
Consulting or Advisory Role: Myriad Genetics
Speakers' Bureau: AstraZeneca

Matthew Rabinowitz
Employment: Natera, MyOme, Marble Therapeutics
Leadership: Natera, MyOme, Marble Therapeutics
Stock and Other Ownership Interests: Natera, MyOme, Marble Therapeutics
Consulting or Advisory Role: Natera, MyOme, Marble Therapeutics
Research Funding: Natera, MyOme
Patents, Royalties, Other Intellectual Property: Natera, MyOme
Travel, Accommodations, Expenses: Natera, MyOme

Akash Kumar
Employment: MyOme
Leadership: MyOme
Stock and Other Ownership Interests: MyOme, Invitae, Sema4
Research Funding: MyOme
Patents, Royalties, Other Intellectual Property: One patent licensed to Illumina, Several patents in the course of work at MyOme

Kate M. Im
Employment: MyOme
Stock and Other Ownership Interests: MyOme
Patents, Royalties, Other Intellectual Property: Several patents in the course of work at MyOme

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

This research was conducted using the UKB resource (project 48991). This study makes use of data generated by the UK10K Consortium, derived from samples from EGAS00001000713. A full list of the investigators who contributed to the generation of the data is available in

www.UK10K.org. The GWAS of Breast Cancer in the African Diaspora is conducted by the University of Chicago and supported by the National Cancer Institute (R01 CA142996-02). Assistance with phenotype harmonization, SNP selection, data cleaning, meta-analyses, data management and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01). Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GARNET Coordinating Center (U01 HG005157). Assistance with data cleaning was provided by the National Center for Biotechnology Information.

REFERENCES

1. Engel C, Fischer C: Breast cancer risks and risk prediction models. Breast Care 10:7-12, 2015
2. Lee A, Mavaddat N, Wilcox AN, et al: BOADICEA: A comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med 21:1708-1718, 2019
3. Antoniou A, Pharoah PDP, Narod S, et al: Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: A combined analysis of 22 studies. Am J Hum Genet 72:1117-1130, 2003
4. Antoniou AC, Spurdle AB, Sinilnikova OM, et al: Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. Am J Hum Genet 82:937-948, 2008
5. Mavaddat N, Antoniou AC, Easton DF, et al: Genetic susceptibility to breast cancer. Mol Oncol 4:174-191, 2010
6. Bogdanova N, Helbig S, Dörk T: Hereditary breast cancer: Ever more pieces to the polygenic puzzle. Hered Cancer Clin Pract 11:12, 2013
7. Couch FJ, Shimelis H, Hu C, et al: Associations between cancer predisposition testing panel genes and breast cancer. JAMA Oncol 3:1190-1196, 2017
8. NCCN: Genetic/Familial High-Risk Assessment: Breast, Ovarian and Pancreatic (Version 2.2022). NCCN, 2022. https://www.nccn.org/guidelines/guidelines-detail?category=2&id=1503
9. Breast Cancer Association Consortium, Dorling L, Carvalho S, et al: Breast cancer risk genes - association analysis in more than 113,000 women. N Engl J Med 384:428-439, 2021
10. Hall P, Easton D: Breast cancer screening: Time to target women at risk. Br J Cancer 108:2202-2204, 2013
11. Michailidou K, Hall P, Gonzalez-Neira A, et al: Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 45:353-361, 361e1-2, 2013
12. Mavaddat N, Pharoah PDP, Michailidou K, et al: Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst 107:djv036, 2015
13. Michailidou K, Beesley J, Lindstrom S, et al: Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet 47:373-380, 2015
14. Michailidou K, Lindström S, Dennis J, et al: Association analysis identifies 65 new breast cancer risk loci. Nature 551:92-94, 2017
15. Milne RL, Kuchenbaecker KB, Michailidou K, et al: Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet 49:1767-1778, 2017
16. van den Broek JJ, Schechter CB, van Ravesteyn NT, et al: Personalizing breast cancer screening based on polygenic risk and family history. J Natl Cancer Inst 113:434-442, 2021
17. Wolfson M, Gribble S, Pashayan N, et al: Potential of polygenic risk scores for improving population estimates of women's breast cancer genetic risks. Genet Med 23:2114-2121, 2021
18. Mealiffe ME, Stokowski RP, Rhees BK, et al: Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. J Natl Cancer Inst 102:1618-1627, 2010
19. Lakeman IMM, Hilbers FS, Rodríguez-Girondo M, et al: Addition of a 161-SNP polygenic risk score to family history-based risk prediction: Impact on clinical management in non-BRCA1/2 breast cancer families. J Med Genet 56:581-589, 2019
20. Hughes E, Tshiaba P, Wagner S, et al: Integrating clinical and polygenic factors to predict breast cancer risk in women undergoing genetic testing. JCO Precis Oncol 5:307-316, 2021
21. 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 25:359-365, 2016
22. 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 159:513-525, 2016
23. Yanes T, Young M-A, Meiser B, et al: Clinical applications of polygenic breast cancer risk: A critical review and perspectives of an emerging field. Breast Cancer Res 2221:21, 2020
24. Kapoor PM, Mavaddat N, Choudhury PP, et al: Combined associations of a polygenic risk score and classical risk factors with breast cancer risk. J Natl Cancer Inst 113:329-337, 2021
25. Li SX, Milne RL, Nguyen-Dumont T, et al: Prospective evaluation of the addition of polygenic risk scores to breast cancer risk models. JNCI Cancer Spectr 5:pkab021, 2021
26. Pharoah PDP, Antoniou A, Bobrow M, et al: Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 31:33-36, 2002
27. 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 17:147, 2015
28. Cuzick J, Sestak I, Cawthorn S, et al: Tamoxifen for prevention of breast cancer: Extended long-term follow-up of the IBIS-I breast cancer prevention trial. Lancet Oncol 16:67-75, 2015
29. 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 35:743-750, 2017
30. 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 54:111-113, 2017
31. 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 4:476-482, 2018
32. Mavaddat N, Michailidou K, Dennis J, et al: Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 104:21-34, 2019
33. Hughes E, Tshiaba P, Gallagher S, et al: Development and validation of a clinical polygenic risk score to predict breast cancer risk. JCO Precis Oncol 4:585-592, 2020

34. Liu C, Zeinomar N, Chung WK, et al: Generalizability of polygenic risk scores for breast cancer among women with European, African, and Latinx ancestry. JAMA Netw Open 4:e2119084, 2021
35. Wang S, Qian F, Zheng Y, et al: Genetic variants demonstrating flip-flop phenomenon and breast cancer risk prediction among women of African ancestry. Breast Cancer Res Treat 168:703-712, 2018
36. Lewis CM, Vassos E: Polygenic risk scores: From research tools to clinical instruments. Genome Med 12:44, 2020
37. Tshiaba P, Sun J, Ratman D, et al: Cross-ancestry polygenic risk score for breast cancer risk assessment. J Clin Oncol 40, 2022 (suppl 16; abstr 10540)
38. UK Biobank: UK Biobank. https://www.ukbiobank.ac.uk/
39. Women's Health Initiative: Women's Health Initiative (WHI). Women's Health Initiative. https://www.whi.org/
40. Chen F, Chen GK, Stram DO, et al: A genome-wide association study of breast cancer in women of African ancestry. Hum Genet 132:39-48, 2013
41. The Multiethnic Cohort Study. https://www.uhcancercenter.org/for-researchers/mec-cohort-composition
42. Howie B, Fuchsberger C, Stephens M, et al: Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44:955-959, 2012
43. Delaneau O, Marchini J; The 1000 Genomes Project Consortium: Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun 5:3934, 2014
44. Tyrer J, Duffy SW, Cuzick J: A breast cancer prediction model incorporating familial and personal risk factors. Stat Med 23:1111-1130, 2004
45. US Preventive Services Task Force, Owens DK, Davidson KW, et al: Medication use to reduce risk of breast cancer: US Preventive Services Task Force recommendation statement. JAMA 322:857-867, 2019
46. Mars N, Lindbohm JV, Briotta Parolo PD, et al: Systematic comparison of family history and polygenic risk across 24 common diseases. bioRxiv 10.1101/2022.07.06.22277333
47. Zhang X, Rice M, Tworoger SS, et al: Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. PLoS Med 15:e1002644, 2018
48. Pal Choudhury P, Wilcox AN, Brook MN, et al: Comparative validation of breast cancer risk prediction models and projections for future risk stratification. J Natl Cancer Inst 112:278-285, 2020
49. Hurson AN, Pal Choudhury P, Gao C, et al: Prospective evaluation of a breast-cancer risk model integrating classical risk factors and polygenic risk in 15 cohorts from six countries. Int J Epidemiol 50:1897-1911, 2022
50. Oeffinger KC, Fontham ETH, Etzioni R, et al: Breast cancer screening for women at average risk: 2015 Guideline update from the American Cancer Society. JAMA 314:1599-1614, 2015