Breast Cancer Research

SHORT REPORT

Open Access

Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry

Parichoy Pal Choudhury[1], Mark N. Brook[2], Amber N. Hurson[1,3], Andrew Lee[4], Charlotta V. Mulder[1], Penny Coulson[2], Minouk J. Schoemaker[2], Michael E. Jones[2], Anthony J. Swerdlow[2,5], Nilanjan Chatterjee[6], Antonis C. Antoniou[4] and Montserrat Garcia-Closas[1]*

Abstract

Background: The Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) and the Tyrer-Cuzick breast cancer risk prediction models are commonly used in clinical practice and have recently been extended to include polygenic risk scores (PRS). In addition, BOADICEA has also been extended to include reproductive and lifestyle factors, which were already part of Tyrer-Cuzick model. We conducted a comparative prospective validation of these models after incorporating the recently developed 313-variant PRS.

Methods: Calibration and discrimination of 5-year absolute risk was assessed in a nested case-control sample of 1337 women of European ancestry (619 incident breast cancer cases) aged 23–75 years from the Generations Study.

Results: The extended BOADICEA model with reproductive/lifestyle factors and PRS was well calibrated across risk deciles; expected-to-observed ratio (*E/O*) at the highest risk decile :0.97 (95 % CI 0.51 − 1.86) for women younger than 50 years and 1.09 (0.66 − 1.80) for women 50 years or older. Adding reproductive/lifestyle factors and PRS to the BOADICEA model improved discrimination modestly in younger women (area under the curve (AUC) 69.7 % vs. 69.1%) and substantially in older women (AUC 64.6 % vs. 56.8%). The Tyrer-Cuzick model with PRS showed evidence of overestimation at the highest risk decile: *E/O* = 1.54(0.81 − 2.92) for younger and 1.73 (1.03 − 2.90) for older women.

Conclusion: The extended BOADICEA model identified women in a European-ancestry population at elevated breast cancer risk more accurately than the Tyrer-Cuzick model with PRS. With the increasing availability of PRS, these analyses can inform choice of risk models incorporating PRS for risk stratified breast cancer prevention among women of European ancestry.

Keywords: Absolute risk, BOADICEA, Breast cancer, IBIS, Model validation, Prospective cohort, Risk prediction, Tyrer-Cuzick

* Correspondence: montserrat.garcia-closas@nih.gov
[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute of Health, 9609 Medical Center Drive 7E-342, Rockville, MD 20850, USA
Full list of author information is available at the end of the article

Introduction

The Breast and Ovarian Analysis of Disease and Carrier Estimation Algorithm (BOADICEA) breast cancer model was originally developed to predict breast cancer risk for women using pedigree-level family history information and genetic testing results on rare pathogenic variants in high and moderate risk genes [1, 2]. This model has been updated (version 5.0) to include reproductive and lifestyle factors and the recently developed polygenic risk score (PRS) based on 313 common germline variants [3] for applications in both general and high-risk populations [4, 5]. The Tyrer-Cuzick or International Breast Intervention Study (IBIS) model [6], commonly used in clinical and research settings, also includes extensive family history and comprehensive risk factor information and has been updated to include information on PRS. We recently evaluated the performance of the Tyrer-Cuzick model (version 8.0) without PRS in a prospective cohort of women of European ancestry in the general population [7]. Here, we perform a comparative validation of the extended BOADICEA and Tyrer-Cuzick models incorporating the 313-variant PRS in the same prospective cohort.

The original BOADICEA and Tyrer-Cuzick models are considered in clinical guidelines [8, 9] for management of women with a family history of breast cancer and have been implemented in user-friendly risk assessment tools that can incorporate PRS (BOADICEA: https://canrisk.org; IBIS: https://ibis.ikonopedia.com). Assessment of disease risk using PRS has been increasingly commercially available through genetic services and marketed to clinicians [10]. Given the widespread use of these models and their capabilities to incorporate PRS with other risk factors in flexible risk prediction tools, comparative prospective validation of these models and their extensions is critical to assess their ability to accurately identify women at different risk levels for risk-stratified screening, surveillance or prevention strategies.

We report results from a prospective comparative validation of the extended BOADICEA model (v.5) with risk factors and 313-variant PRS and Tyrer-Cuzick model (v.8) with the same PRS in the Generations Study, a population-based cohort study of UK women [11].

Methods

Data were used from a nested case-control sample within the Generations Study (2003–2012), a prospective cohort of over 113,000 UK women aged 16–102 years; details are elsewhere [7, 11]. The comparative validation analyses of 5-year absolute risk of breast cancer were based on 1337 women aged 23–75 years, including 619 incident breast cancer patients within 5 years from study recruitment, with information on the PRS and the risk factors used in both the BOADICEA (v.5) and Tyrer-Cuzick (v.8) models (Supplementary Fig. 1). Supplementary Table 1 summarizes the information on questionnaire-based risk factors and 313-variant PRS for these women.

To update the original BOADICEA model, the relative risks for the risk factors and PRS were derived using the literature-based approach [3, 7]; further details are given in Lee et al. [4]. In this model, the family history association, described by a residual polygenic component, was adjusted to account for the PRS explaining ~ 20% of the breast cancer familial aggregation. The PRS was added to the Tyrer-Cuzick model (v.8) using the approach described in Brentnall et al. [12], where the associations of family history and PRS were unadjusted and assumed to be multiplicative on the risk of developing breast cancer. The comparative validation analyses were performed using the standardized model calibration and discrimination methods implemented in the Individualized Coherent Absolute Risk Estimator (iCARE) tool [13] (details in supplement). Briefly, model calibration was assessed in terms of relative and absolute risk by comparing the observed and expected quantities, overall, and within risk categories. The area under the curve (AUC) was estimated to assess model discrimination.

Results

For women younger than 50 years, the original and extended BOADICEA models (with PRS and with PRS and reproductive/lifestyle factors) showed good calibration of relative and absolute risk (Fig. 1). At the highest decile of predicted 5-year absolute risk, the extended model with PRS and reproductive/lifestyle factors showed better calibration than both the original model and the extended model with PRS only, with a ratio of expected to observed number of cases (E/O) of 0.97 [95% confidence interval (CI) 0.51 – 1.86], 0.83 (0.44 – 1.56), 0.85 (0.44 – 1.63), respectively. Adding PRS and risk factors led to modest improvement in AUC from 69.1 % (63.5 % – 74.6%) to 69.7 % (64.1 % – 75.2%). Incorporating risk factors did not improve the discrimination of the original model (data not shown) or the extended model with PRS (Fig. 1). The Tyrer-Cuzick model with PRS had similar discrimination [AUC : 69.4 % (63.8 % – 75.0%)] to the extended BOADICEA model with PRS and risk factors but showed evidence of overestimation at the highest risk decile [E/O : 1.54 (0.81 – 2.92)].

The original and extended BOADICEA models also showed good calibration of relative and absolute risk for women 50 years or older (Fig. 2), in particular for women at the highest risk decile [E/O : 0.95 (0.56 – 1.62) for the original model, 1.07 (0.63 – 1.82) for the extended model with PRS, 1.09 (0.66 – 1.80) for the extended model with PRS and risk factors]. For this age group, incorporating PRS and risk factors led to

Pal Choudhury *et al. Breast Cancer Research*   (2021) 23:22
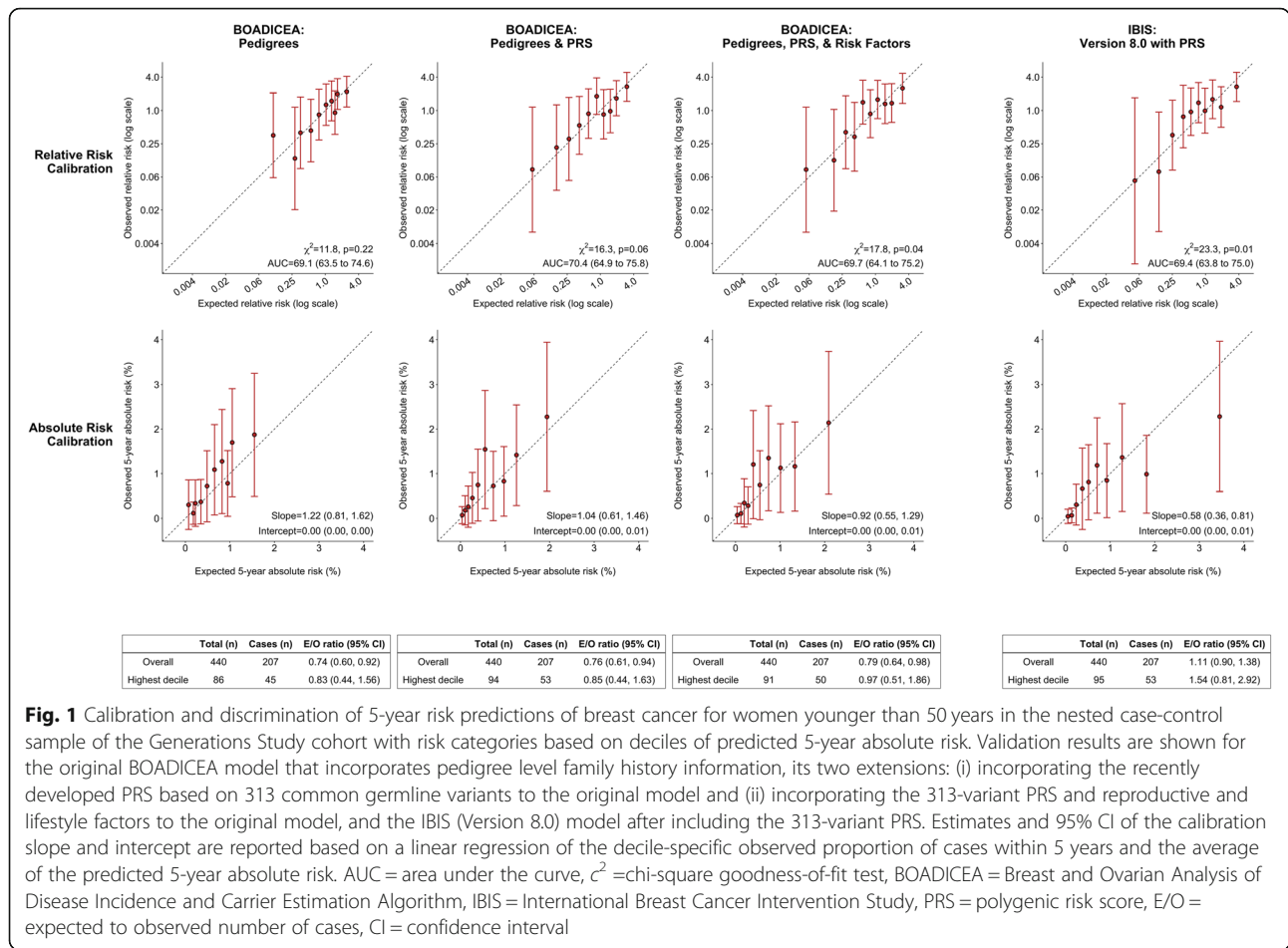
Page 3 of 5

Fig. 1 Calibration and discrimination of 5-year risk predictions of breast cancer for women younger than 50 years in the nested case-control sample of the Generations Study cohort with risk categories based on deciles of predicted 5-year absolute risk. Validation results are shown for the original BOADICEA model that incorporates pedigree level family history information, its two extensions: (i) incorporating the recently developed PRS based on 313 common germline variants to the original model and (ii) incorporating the 313-variant PRS and reproductive and lifestyle factors to the original model, and the IBIS (Version 8.0) model after including the 313-variant PRS. Estimates and 95% CI of the calibration slope and intercept are reported based on a linear regression of the decile-specific observed proportion of cases within 5 years and the average of the predicted 5-year absolute risk. AUC = area under the curve, c^2 =chi-square goodness-of-fit test, BOADICEA = Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm, IBIS = International Breast Cancer Intervention Study, PRS = polygenic risk score, E/O = expected to observed number of cases, CI = confidence interval

substantial improvements in AUC from 56.8 % (52.9 % – 60.6%) to 64.6 % (60.9 % – 68.2%). Adding risk factors substantially improved the risk discrimination of the original model (data not shown) and the extended model with PRS (Fig. 2). The Tyrer-Cuzick model with PRS had risk discrimination comparable to the extended BOADICEA model with PRS and risk factors; however, the former substantially overestimated risk for women at the highest risk decile [E/O : 1.73 (1.03 – 2.90)]. Overestimation of risk in high-risk deciles was present in models with or without the PRS (Supplementary Fig. 2).

Discussion

Our study shows that the extended BOADICEA model, which incorporated reproductive and lifestyle factors and a 313-SNP PRS to the familial aggregation information, predicted 5-year absolute risk of breast cancer more accurately than the Tyrer-Cuzick model with the same PRS, for women at the highest risk decile in the Generations study, a UK-based prospective cohort.

Previous studies in populations of women of European ancestry provided evidence of overestimation of absolute risk obtained from the Tyrer-Cuzick model without PRS

for women in the highest risk decile [7, 14]. Two recent studies that incorporated PRSs with fewer genetic variants to this model, showed good calibration in terms of relative risk, but did not evaluate absolute risk calibration [12, 15]. Our results showed overestimation of the absolute risk for women at the high-risk deciles, possibly due to not attenuating the contribution of family history association to account for the substantial familial aggregation explained by the PRS. This can lead to inflated breast cancer risks, particularly for women with breast cancer family history who are more prevalent in high-risk deciles. Accounting for the correlation between the PRS and family history would likely reduce this overestimation and future studies are needed to investigate the extent of this reduction.

Strengths of the current analyses include the use of the Generations Study, a relatively recent population-based cohort with a wide range of ages of participating women and the comparison of two widely used risk prediction tools that can incorporate PRS. With the increasing availability of PRS (e.g., in countries like US), such rigorous comparative evaluation of models incorporating PRS with other risk factor information is critical to
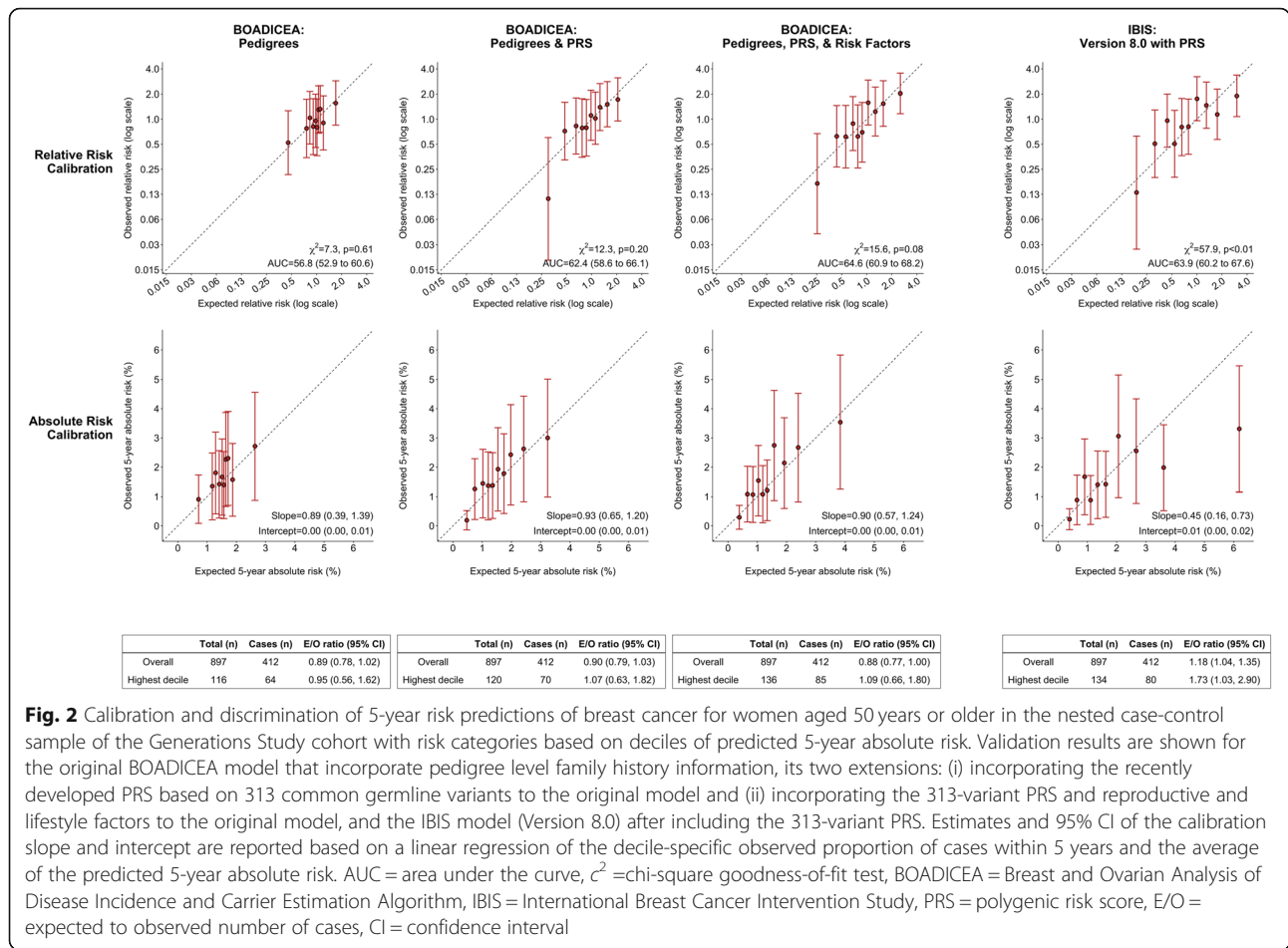
Pal Choudhury *et al. Breast Cancer Research*   (2021) 23:22

Page 4 of 5

BOADICEA: Pedigrees

	Total (n)	Cases (n)	E/O ratio (95% CI)
Overall	897	412	0.89 (0.78, 1.02)
Highest decile	116	64	0.95 (0.56, 1.62)

BOADICEA: Pedigrees & PRS

	Total (n)	Cases (n)	E/O ratio (95% CI)
Overall	897	412	0.90 (0.79, 1.03)
Highest decile	120	70	1.07 (0.63, 1.82)

BOADICEA: Pedigrees, PRS, & Risk Factors

	Total (n)	Cases (n)	E/O ratio (95% CI)
Overall	897	412	0.88 (0.77, 1.00)
Highest decile	136	85	1.09 (0.66, 1.80)

IBIS: Version 8.0 with PRS

	Total (n)	Cases (n)	E/O ratio (95% CI)
Overall	897	412	1.18 (1.04, 1.35)
Highest decile	134	80	1.73 (1.03, 2.90)

Fig. 2 Calibration and discrimination of 5-year risk predictions of breast cancer for women aged 50 years or older in the nested case-control sample of the Generations Study cohort with risk categories based on deciles of predicted 5-year absolute risk. Validation results are shown for the original BOADICEA model that incorporate pedigree level family history information, its two extensions: (i) incorporating the recently developed PRS based on 313 common germline variants to the original model and (ii) incorporating the 313-variant PRS and reproductive and lifestyle factors to the original model, and the IBIS model (Version 8.0) after including the 313-variant PRS. Estimates and 95% CI of the calibration slope and intercept are reported based on a linear regression of the decile-specific observed proportion of cases within 5 years and the average of the predicted 5-year absolute risk. AUC = area under the curve, c^2 = chi-square goodness-of-fit test, BOADICEA = Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm, IBIS = International Breast Cancer Intervention Study, PRS = polygenic risk score, E/O = expected to observed number of cases, CI = confidence interval

assess their suitability in clinical and research applications. Moreover, model calibration was assessed both overall and within risk categories, in particular for women at the extremes of risk for whom prevention and screening are most relevant. The CanRisk tool has already implemented the BOADICEA model and its extensions. The current study provides some evidence of accurate risk predictions from this tool for the UK general population. Further evaluation of this tool in both general and high-risk populations is needed before widespread clinical applications. Moreover, future research is merited towards risk model building and validation for women of non-European ancestry.

To summarize, the extended BOADICEA model with PRS and reproductive/lifestyle factors identified women of European ancestry at elevated 5-year risk of breast cancer more accurately than the Tyrer-Cuzick model with PRS. As disease risk prediction with PRS is becoming more available through genetic services in some countries (e.g., the USA), these and other similar analyses will potentially inform the choice of risk models for developing risk-stratified breast cancer prevention and screening strategies for women of European ancestry.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13058-021-01399-7.

Additional file 1: Supplementary Materials: Comparative validation of the BOADICEA and Tyrer- Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry.Additional details on the definition of study follow-up, sources of genotype data, risk factors in BOADICEA and Tyrer-Cuzick models, model validation methods are given. **Supplementary Fig. 1**. shows the study design in the validation cohort. **Supplementary Fig. 2**. shows a comparative validation of Tyrer-Cuzick model with and without PRS in the Generations Study. **Supplementary Table 1** shows the risk factor distribution in the Generation Study. Figure 1. Calibration and discrimination of five-year risk predictions of breast cancer for women younger than 50 years in the nested case-control sample of the Generations Study cohort with risk categories based on deciles of predicted five-year absolute risk. Figure 2. Calibration and discrimination of five-year risk predictions of breast cancer for women aged 50 years or older in the nested case-control sample of the Generations Study cohort with risk categories based on deciles of predicted five-year absolute risk.

Pal Choudhury *et al. Breast Cancer Research*   (2021) 23:22

Page 5 of 5

Availability of data and materials
All data relevant to this report are included in this published article and its supplementary information files. The access policy for analysis of the Generations study datasets is available at: https://www.breakthroughgenerations.org.uk/researchers.

Ethics approval and consent to participate
The Generations Study was approved under the procedures for national medical research studies, by the South-East Multi-Centre Research Ethics Committee.

Consent for publication
Not applicable.

Competing interests
Breast and Ovarian Analysis of Disease and Carrier Estimation Algorithm (BOADICEA) has been licensed to Cambridge Enterprise (University of Cambridge) and Antonis C. Antoniou and Andrew Lee are listed as creators. No other disclosures are reported.

Author details
[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute of Health, 9609 Medical Center Drive 7E-342, Rockville, MD 20850, USA. [2]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. [3]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [4]Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. [5]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK. [6]Department of Biostatistics, The Johns Hopkins University, MD, Baltimore, USA.

References
1. Antoniou AC, Pharoah PPD, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br J Cancer. 2004. https://doi.org/10.1038/sj.bjc.6602175.
2. Lee AJ, Cunningham AP, Tischkowitz M, et al. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. Genet Med. 2016;18(12):1190–8. https://doi.org/10.1038/gim.2016.31.
3. 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. https://doi.org/10.1016/j.ajhg.2018.11.002.
4. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med. 2019. https://doi.org/10.1038/s41436-018-0406-9.
5. Lakeman IMM, Rodríguez-Girondo M, Lee A, et al. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. Genet Med. 2020. https://doi.org/10.1038/s41436-020-0884-4.
6. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004. https://doi.org/10.1002/sim.1668.
7. Choudhury PP, Wilcox AN, Brook MN, et al. Comparative validation of breast cancer risk prediction models and projections for future risk stratification. JNCI J Natl Cancer Inst. 2019. https://doi.org/10.1093/jnci/djz113.
8. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®): Breast Cancer Risk Reduction Version 1.2019. https://www.nccn.org/professionals/physician_gls/pdf/breast_risk.pdf. Published 2018.
9. NICE. Familial breast cancer: classification , care and managing breast cancer and related risks in people with a family history of breast cancer. Natl institue Heal Care Excell UK. 2013. https://doi.org/10.1186/1471-2105-9-439.
10. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19(9):581–90. https://doi.org/10.1038/s41576-018-0018-x.
11. Swerdlow AJ, Jones ME, Schoemaker MJ, et al. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. Br J Cancer. 2011. https://doi.org/10.1038/bjc.2011.337.
12. Brentnall AR, van Veen EM, Harkness EF, et al. A case-control evaluation of 143 single nucleotide polymorphisms for breast cancer risk stratification with classical factors and mammographic density. Int J Cancer. 2019. https://doi.org/10.1002/ijc.32541.
13. Pal Choudhury P, Maas P, Wilcox A, et al. iCARE: an R package to build, validate and apply absolute risk models. PLoS One. 2020;15(2):e0228198. https://doi.org/10.1371/journal.pone.0228198.
14. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term accuracy of breast cancer risk assessment combining classic risk factors and breast density. JAMA Oncol. 2018. https://doi.org/10.1001/jamaoncol.2018.0174.
15. 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. https://doi.org/10.1001/jamaoncol.2017.4881.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.