Incorporating Alternative Polygenic Risk Scores into the BOADICEA Breast Cancer Risk Prediction Model

Nasim Mavaddat[1], Lorenzo Ficorella[1], Tim Carver[1], Andrew Lee[1], Alex P. Cunningham[1], Michael Lush[1], Joe Dennis[1], Marc Tischkowitz[2], Kate Downes[3], Donglei Hu[4], Eric Hahnen[5,6], Rita K. Schmutzler[5,6,7], Tracy L. Stockley[8,9,10], Gregory S. Downs[8,10], Tong Zhang[8], Anna M. Chiarelli[11,12], Stig E. Bojesen[13,14,15], Cong Liu[16], Wendy K. Chung[17], Monica Pardo[18], Lidia Feliubadaló[19,20,21], Judith Balmaña[18,22], Jacques Simard[23], Antonis C. Antoniou[1], and Douglas F. Easton[1,24]

ABSTRACT

Background: The multifactorial risk prediction model BOADI-CEA enables identification of women at higher or lower risk of developing breast cancer. BOADICEA models genetic susceptibility in terms of the effects of rare variants in breast cancer susceptibility genes and a polygenic component, decomposed into an unmeasured and a measured component - the polygenic risk score (PRS). The current version was developed using a 313 SNP PRS. Here, we evaluated approaches to incorporating this PRS and alternative PRS in BOADICEA.

Methods: The mean, SD, and proportion of the overall polygenic component explained by the PRS (α^2) need to be estimated. α was estimated using logistic regression, where the age-specific log-OR is constrained to be a function of the age-dependent polygenic relative risk in BOADICEA; and using a retrospective likelihood (RL)

approach that models, in addition, the unmeasured polygenic component.

Results: Parameters were computed for 11 PRS, including 6 variations of the 313 SNP PRS used in clinical trials and implementation studies. The logistic regression approach underestimates α, as compared with the RL estimates. The RL α estimates were very close to those obtained by assuming proportionality to the OR per 1 SD, with the constant of proportionality estimated using the 313 SNP PRS. Small variations in the SNPs included in the PRS can lead to large differences in the mean.

Conclusions: BOADICEA can be readily adapted to different PRS in a manner that maintains consistency of the model.

Impact : The methods described facilitate comprehensive breast cancer risk assessment.

Introduction

BOADICEA (1, 2) is a risk prediction algorithm for predicting breast and ovarian cancer risk on the basis of genetic and nongenetic factors. The algorithm incorporates the effects of common genetic variants, summarized in a polygenic risk score (PRS), in addition to the effects of pathogenic variants in major breast cancer susceptibility genes, other lifestyle/hormonal risk factors, and cancer family history.

The current version (v6; refs. 1, 2) has been specifically developed to incorporate the 313 SNP PRS of Mavaddat and colleagues (3); this PRS was developed using the very large data set of the Breast Cancer Association Consortium (BCAC) and extensively validated in prospective studies. However, as larger genome-wide association studies (GWAS) and novel statistical methods become available, new PRS are being continually developed. In addition, PRS developed for clinical

[1]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom. [2]Department of Medical Genetics and National Institute for Health Research, Cambridge Biomedical Research Centre, The University of Cambridge, Cambridge, United Kingdom. [3]Cambridge Genomics Laboratory, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom. [4]Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, California. [5]Center for Familial Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [6]Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [7]Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [8]Advanced Molecular Diagnostics Laboratory, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [9]Department of Laboratory Medicine and Pathobiology, The University of Toronto, Ontario, Canada. [10]Division of Clinical Laboratory Genetics, Laboratory Medicine Program, University Health Network, Toronto, Canada. [11]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. [12]Ontario Health, Cancer Care Ontario, Toronto, Ontario, Canada. [13]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. [14]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. [15]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [16]Department of Biomedical Informatics, Columbia University Irving Medical

Center, New York, New York. [17]Departments of Pediatrics and Medicine, Columbia University, New York, New York. [18]Hereditary Cancer Genetics Group, Vall d'Hebron Institut d'Oncologia, Barcelona, Spain. [19]Hereditary Cancer Program, Catalan Institute of Oncology (ICO), L'Hospitalet de Llobregat, Spain. [20]Program in Molecular Mechanisms and Experimental Therapy in Oncology (Oncobell), IDIBELL, L'Hospitalet de Llobregat, Spain. [21]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain. [22]Medical Oncology Department, University Hospital of Vall d'Hebron, Barcelona, Spain. [23]Department of Molecular Medicine, Université Laval and CHU de Québec-Université Laval Research Center, Québec, Canada. [24]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom.

Corresponding Author: Nasim Mavaddat, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts' Causeway, Cambridge CB1 8RN, United Kingdom. Phone: 44-1223-748629; Fax: 44-1223-748628; E-mail: nm274@medschl.cam.ac.uk

translation and generated in different health care systems use a variety of technologies, including both targeted sequencing panels and genotyping arrays, and surrogate SNPs are often required. The BOADICEA algorithm itself is flexible and can incorporate any PRS for which the relevant parameters are known. These parameters are the mean (μ) and SD (σ) of the PRS in the population, and the proportion (α^2) of the polygenic variance attributable to the PRS. In practice, the PRS can be normalized and supplied as a Z-score, in which case only parameter α is required. By modelling the PRS as the proportion of a (fixed) polygenic component, the predicted familial risks remain consistent, irrespective of the PRS used, and importantly, there is no double counting of the effect of the PRS and cancer family history.

Here, we discuss the incorporation of alternative PRSs into BOADICEA, and provide the relevant parameters for a number of PRS that have been developed, including several that are in use in clinical applications.

Materials and Methods

BOADICEA models breast cancer risks such that the incidence of breast cancer at age t is of the form (1, 4):

$$\lambda(t) \ = \ \lambda_0(t)exp(\delta_{g(i)}(t) + \sigma_P(t)x_P^{(i)} + \sum_\rho \beta_\rho z_{\rho i}) \qquad (A)$$

Here $\lambda_0(t)$ is the baseline incidence. The term $\delta_{g(i)}(t)$ models the major gene component for individual i ($\delta_k(t)$ being the age-specific log-HR associated with genotype k). $\sigma_P(t)x_P^{(i)}$ models the polygenic component, $\sigma_P(t)$ being the polygenic SD and $x_P^{(i)}$ the normalized polygenic component for individual i. The final term models the effects of other risk factors. The polygenic variance $\sigma_P^2(t)$ is allowed to be age-dependent and assumed to be a linear function of age t:

$$\sigma_P^2(t) \ = \ \gamma + \theta t$$

The parameters γ and θ have been previously estimated, using complex segregation analysis, as 4.86 and -0.06 respectively (4).

The PRS is incorporated into BOADICEA by partitioning the total polygenic component x_P into the sum of a known component x_K measured by the PRS, and an unmeasured residual component x_R (1). The variance due to the known component is of the form (3):

$$\sigma_K^2(t) \ = \ \alpha^2(\gamma + \theta t) \qquad (B)$$

$\sigma_K(t)$ can also be interpreted as the age-specific log-HR per unit SD of the PRS, conditional on other risk factors. Note that in Mavaddat and colleagues (3) equation (B) is written $\sigma_K^2(t) \ = \ \gamma^2(\alpha + \beta t)$. The change of symbols is for consistency with Lee and colleagues (1) and the Canrisk platform (www.canrisk.org), where the proportion of the polygenic variance explained by the PRS is denoted as α^2.

Estimation of α and incorporating alternative PRS

The key parameter is α. The first approach to estimating this parameter makes the simplifying assumption that the polygenic SD of the known polygenic component in BOADICEA, $\sigma_K(t)$ can be approximated by the marginal age-specific log-HR per unit SD of the PRS (ref. 3; see Supplementary Methods). This can then be estimated using cohort data or (approximately, making the rare disease assumption) case–control data, by first transforming the PRS using:

$$S' \ = \ x_K\sqrt{\gamma + \theta t} \qquad (C)$$

where x_K is the standardised (per unit SD) version of the proposed PRS. S' is then included as a covariate in a Cox or logistic regression model: the parameter (log-HR or log-OR) estimate corresponding to the covariate S' gives the required α parameter, which we denote α_{GLM}. This method was applied to 22,767 controls and 16,151 women diagnosed with invasive breast cancer from the validation and prospective test sets used in Mavaddat and colleagues (ref. 3; Supplementary Tables S1 and S2). The analysis was restricted to women of European ancestry with age of diagnosis or last observation less than 80 years [after application of inclusion/exclusion criteria, mean age at diagnosis = 59.9 (SD = 10) years for cases, and 57.1 (SD = 10.4) years for controls]. Analyses were adjusted for country in which the study was conducted (15 countries) and 10 principal components.

The above analyses make the simplifying assumption that the marginal PRS effect size is a good approximation to the effect size conditional on other risk factors. This is likely to be a reasonable assumption for nongenetic risk factors, which have relatively small effects on risk and appear to be independent of the PRS, as shown in recent analyses of the combined effect of breast cancer PRS and individual SNPs with life-style/environmental risk factors including questionnaire-based factors (5–10). However, it may not be true for other genetic factors, in particular the unmeasured polygenic component. Although the PRS and the residual polygenic component are assumed to be conditionally independent, individuals with a high polygenic component are more likely to develop the disease at an early age. This results in a negative correlation between the PRS and the residual polygenic component at later ages, which leads to an underestimation of the PRS effect size if the latter is not allowed for. To address this problem, we also estimated α using a retrospective likelihood approach (α_{RL}), applied to the same BCAC data set. In this analysis, the observed PRS is computed conditional on the phenotypes of the individuals (age of diagnosis and case–control status), explicitly allowing for the unmeasured polygenic component. Details are given in the Supplementary Methods. This approach requires overall population age-specific incidence rates to be specified. For this purpose, the rates for England and Wales 2016–2018 were used (https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/age).

Because the mean PRS varies by country, we first regressed the PRS on country and principal components, adjusted for case–control status, and performed the analyses on the residual PRS. The likelihood was maximized using the optimize function in R. 95% confidence intervals (CI) were obtained using a grid of values for α_{RL}, and finding the difference between the log-likelihoods and the maximum log-likelihood.

As a third approach, we derived an approximate estimate α from the log-OR per unit SD (η), by calibrating against PRS$_{313}$ as a standard. From equations (A) and (B) in the methods above it can be seen that, under the rare disease assumption, the marginal HR associated with the PRS should approximate the conditional HR. If differential age effects can also be ignored, α should therefore be approximately proportional to η. This allows α to be estimated using PRS313 as a standard. Thus: $\alpha_{APP} \ = \ \frac{\eta}{\eta_0}\alpha_0$ where η_0 and α_0 are the corresponding estimates for PRS313. This provides a simple method that could be applied to PRS developed and validated on a different data set.

We computed the relevant parameters for PRS313 and 10 additional PRS [Supplementary Tables S3 and S4; SNP positions based on Genome Reference Consortium Human Build 37 (GRCh37)].

PRS313 includes two variants (22_29203724_C_T and 22_29551872_A_G) which are correlated with the protein truncating variant *CHEK2**1100delC, and some of the derivative PRS also include these SNPs. This could result in overestimation of risk in *CHEK2**1100delC carriers if the PRS is used in conjunction with gene-panel testing, because BOADICEA assumes that the PRS and major gene genotypes are independent in the population. We therefore also considered PRS without these variants. (Note that *CHEK2* p.I157T (22_29121087_A_G) is also included in PRS313 but is only weakly correlated with *CHEK2**1100delC and does not introduce a bias). The means and SDs of each PRS, and the proportion (α) of the polygenic variance attributable to these alternative PRS were derived in the same data set (Supplementary Tables S1 and S2), namely the validation and prospective sets described by Mavaddat and colleagues (3).

All studies included in this analysis were approved by the relevant local ethical review boards and used appropriate consent procedures. SEARCH was approved by the NRES Committee East of England - Cambridge South.

Data availability

Data were generated by the authors and is available on request.

Results

PRS examples

Eleven alternative PRS were constructed. Six of these are modifications of the PRS313, designed for clinical implementation. The BRIDGES PRS was developed as an next-generation sequencing (NGS) panel test to facilitate clinical translational studies of BOADICEA implemented in the context of genetic testing of women with a family history (https://bridges-research.eu/). Of 313 variants, 295 could be designed and a further 11 were replaced by surrogate markers ($r^2 > 0.9$ in Europeans). The PERSPECTIVE I&I PRS was designed to facilitate risk stratified screening in the context of population-based mammographic screening in Ontario and Quebec (11). This PRS was designed as an NGS panel: 287 of 313 markers could be designed and a further 8 were surrogates. The EastGLH PRS was designed by the NHS East Genomic Laboratory Hub for use in a randomized control trial of women testing positive for an inherited pathogenic/likely pathogenic gene variant in *BRCA1, BRCA2, PALB2, CHEK2,* or *ATM*, using a NGS panel of 303 markers (12). The PRISMA PRS, designed as genotyping array of 268 markers (37 surrogates), was developed to provide multifactorial cancer risks to women attending genetic clinics in Spain. The eMERGE PRS consisted of 308 markers and is part of a large US study aiming to communicate PRS-based genome-informed risk assessment across multiple diseases (https://emerge-network.org). DBDS299, using data from the Danish Blood Donor Study (https://bmjopen.bmj.com/content/9/6/e028401) is used in a research study to validate BOADICEA in the Danish population. In addition, we included the earlier PRS77 developed using BCAC data and comprising genome-wide significant SNPs, PRS3820 developed by Mavaddat and colleagues (3) using Lasso penalized regression, and two PRS (WISDOM75 and WISDOM120) developed for the WISDOM clinical trial (ref. 13; Clinical Trials identifier NCT02620852). We also considered all of the above PRS without 22_29203724_C_T and 22_29551872_A_G, SNPs correlated with *CHEK2**1100delC, as described in Materials and Methods.

PRS parameters

Table 1 summarizes the estimated parameters for PRS313 and each of the alternative PRS. As expected, the 6 PRS that are variations on PRS313 have very similar effect sizes, expressed as log-OR per 1 SD, reflecting the fact that only a few variants are not accounted for. The α_{RL} parameters for these 6 PRS are also similar, and only marginally lower than PRS313 estimate (0.441; 95% CI, 0.430–0.445). The effect sizes for PRS77 (both in terms of the log-OR per 1 SD and α) were smaller than for PRS313, while PRS3820 had larger effect sizes. The two WISDOM and PRISMA PRSs also had somewhat smaller effect sizes than PRS313. Removal of the 2 chromosome 22 SNPs had only a small effect on the estimated log-OR per 1SD, and α – for example reducing α_{RL} from 0.441 to 0.439 for PRS313. The α values computed using the simpler logistic regression approach (α_{GLM}) were smaller than those generated using the α_{RL} for all PRS.

We note that the α_{RL} are approximately proportional to the PRS effect sizes, expressed as odds-ratio per 1 SD (**Table 1**; **Fig. 1**). Using the PRS313 as the standard, the predicted α value assuming proportionality is given by $\alpha_{APP} = 0.887\eta$ (**Table 1**; **Fig. 1**). These predicted values were very similar to the α_{RL} values for all PRS.

Discussion

We evaluated approaches to incorporating alternate breast cancer PRSs into the risk prediction algorithm BOADICEA. The α values computed using the simpler logistic regression approach (α_{GLM}) were consistently smaller than those generated using the α_{RL}, for all PRS. This difference can be explained by the fact that the logistic regression approach does not account for the residual component. Women with a high polygenic component are more likely to develop the disease at an early age, resulting in a negative correlation between the PRS and the residual polygenic component, which leads to an underestimation of the PRS effect size if the latter is not allowed for, a phenomenon related to index event bias (14).

We showed further that the α parameters derived from the log-OR estimate by assuming proportionality were very close to the α_{RL} estimates. This suggests that this approach is likely to be reasonably accurate for other PRS, at least across the range of effect sizes considered here, providing a very straightforward approach to incorporating a PRS developed on another data set if a log-OR estimate is already available.

A striking observation is the very large difference in the means of the different PRS. This reflects the fact that the removal of a few SNPs with important weights can have a substantial effect on the mean. For example, the means for the PRS excluding the chromosome 22 SNPs are higher. While the mean has no intrinsic significance, this emphasizes the importance of correctly normalizing the PRS. In particular, because BOADICEA also incorporates the effects of *CHEK2* protein truncating variants, we recommend using the PRS without these SNPs when gene-panel testing is performed.

It is important to note that estimates derived from European ancestry populations may not be applicable to individuals of other ancestries. The effect sizes may differ among populations, for example due to differences in linkage disequilibrium structure. This has been shown for PRS313, for which somewhat smaller effect sizes have been estimated in Asian and African-American populations (15–18). In addition, the mean PRS can vary significantly by population—PRS313 has a higher mean in both Asian and African-American populations than in Europeans. This again emphasizes the importance of calibrating the PRS to the relevant population distribution. The argument that the α_{RL} is preferable and provides a more reliable estimate of α should also hold in non-European populations.

The analyses used here adjusted the PRS for both the country in which the study was conducted and ancestry informative principal

Table 1. Summary parameters for alternative PRSs.[a]

	No of SNPs	Mean controls	Mean cases	SD controls	SD cases	Log-OR per 1 SD	95% CI	α_{GLM}[b]	95% CI	α_{RL}[c]	95% CI	Ratio (α_{RL}:Log-OR per 1 SD)	Predicted α_{RL}
BCAC PRS313	313	−0.424	−0.114	0.611	0.619	0.497	0.476–0.519	0.397	0.378–0.415	0.441	0.430–0.455	0.887	0.441
BRIDGES	306	−0.422	−0.114	0.608	0.615	0.495	0.474–0.517	0.394	0.376–0.413	0.439	0.425–0.450	0.886	0.439
PERSPECTIVE	295	−0.448	−0.147	0.599	0.609	0.489	0.468–0.511	0.389	0.371–0.408	0.433	0.420–0.450	0.886	0.434
EASTGLH	303	−0.407	−0.100	0.606	0.613	0.494	0.472–0.516	0.394	0.375–0.412	0.438	0.420–0.450	0.886	0.438
PRISMA	268	0.322	0.575	0.558	0.565	0.446	0.424–0.467	0.358	0.339–0.376	0.396	0.385–0.410	0.890	0.395
eMERGE	308	−0.456	−0.150	0.608	0.616	0.495	0.473–0.516	0.394	0.376–0.413	0.439	0.425–0.450	0.887	0.439
DBDS299	299	−0.508	−0.211	0.596	0.605	0.486	0.465–0.508	0.387	0.369–0.406	0.432	0.440–0.455	0.888	0.431
BCAC PRS77	77	−0.892	−0.703	0.449	0.460	0.394	0.373–0.415	0.310	0.292–0.328	0.350	0.340–0.360	0.889	0.349
BCAC PRS3820	3820	−0.445	−0.199	0.460	0.463	0.518	0.496–0.540	0.412	0.393–0.431	0.454	0.440–0.470	0.878	0.459
WISDOM75	74	−1.057	−0.835	0.567	0.573	0.360	0.338–0.381	0.281	0.263–0.299	0.315	0.300–0.330	0.875	0.319
WISDOM128	126	−0.180	0.041	0.464	0.478	0.449	0.427–0.470	0.355	0.337–0.374	0.402	0.390–0.415	0.895	0.398
PRS constructed as above but excluding SNPs correlated with CHEK2*1100delC													
BCAC PRS311	311	−0.092	0.217	0.609	0.618	0.496	0.474–0.518	0.395	0.377–0.414	0.439	0.425–0.455	0.886	0.440
BRIDGES	305	−0.428	−0.121	0.607	0.615	0.495	0.473–0.516	0.394	0.375–0.412	0.438	0.425–0.450	0.886	0.439
PERSPECTIVE	293	−0.116	0.183	0.598	0.607	0.488	0.466–0.510	0.388	0.369–0.406	0.432	0.420–0.445	0.885	0.433
EASTGLH	301	−0.075	0.230	0.604	0.612	0.492	0.471–0.514	0.392	0.374–0.411	0.436	0.420–0.450	0.885	0.437
eMERGE	306	−0.124	0.181	0.606	0.614	0.493	0.471–0.515	0.393	0.374–0.411	0.437	0.425–0.450	0.886	0.437
DBDS299	297	−0.177	0.119	0.594	0.603	0.485	0.463–0.506	0.386	0.367–0.404	0.430		0.887	0.430
BCAC PRS3818	3818	−0.227	0.018	0.459	0.462	0.517	0.495–0.539	0.411	0.392–0.430	0.453	0.440–0.465	0.877	0.458

Abbreviations: GLM, generalized linear model; RL, retrospective likelihood.

[a]Results based on 22767 controls and 16151 cases in the prospective and validation sets from Mavaddat and colleagues (3) . European women with known age and age less than 80.

[b]α_{GLM} is based on logistic regression adjusted for country in which studies were conducted and 10 principal components.

[c]α_{RL} is based on retrospective likelihood method as described in Materials and Methods and Supplementary Methods.

BRIDGES includes 22_29203724_C_T but not 22_29551872_A_G.

BCAC PRS313, BCAC 3820, eMERGE, DBDS299, EASTGLH and PERSPECTIVE include both 22_29203724_C_T and 22_29551872_A_G.

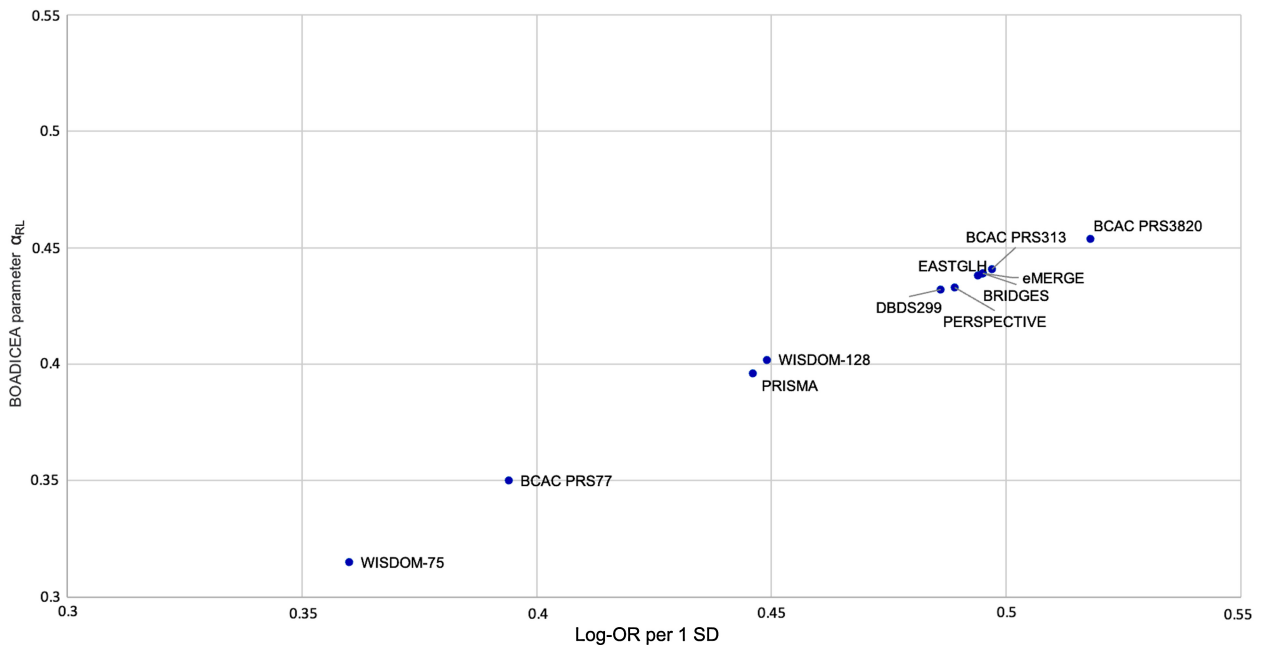PRISMA has neither 22_29551872_A_G nor 22_29203724_C_T and includes 37 surrogates.

Figure 1.
Graph of BOADICEA parameter alpha (α_{RL}) versus log-OR per 1 SD. Graph of BOADICEA parameter alpha (α_{RL}) versus log-OR per 1 SD for PRS313 and alternative PRS. α_{RL} were estimated using the retrospective likelihood method as described in Materials and Methods and Supplementary Methods.

components. An adjustment is necessary because the mean PRS varies by country, even among European populations (and this is not reflected in differences in incidence rates). However, it is possible that adjustment for both country and principal components is over-conservative. Further analyses in large population-specific data sets may be able to address this.

The approaches described allow BOADICEA to be adapted for use with any PRS in a consistent manner. However, it should be emphasized that the main validations of BOADICEA used PRS313 (19–23). For PRS that are substantially different, and particularly as more informative PRS are generated through larger GWAS, further prospective validation in independent external data sets will be required. We also note that the current formulation of BOADICEA assumes that the age-specific effects of the PRS and the residual polygenic component (as measured by the log-HR per 1 SD) are proportional. This significantly simplifies the algorithm, but it is possible that better predictions may be available by allowing differential age-specific effects.

The BOADICEA algorithm has been extensively validated, particularly when incorporating PRS313 (19–23) in addition to other risk factors. It is available through the CanRisk (www.canrisk.org) tool (24) and is widely used in the context of women with cancer family history or undergoing gene-panel testing, including several ongoing clinical implementation studies. The CanRisk tool provides the facility to incorporate a PRS as a Z-score, providing that the α parameter is known. The methods described here allow other PRSs to be used with BOADICEA via CanRisk and hence facilitate more widespread comprehensive breast cancer risk assessment.

Authors' Disclosures

Disclaimer

Authors' Contributions

N. Mavaddat: Conceptualization, formal analysis, writing–original draft, writing–review and editing. **L. Ficorella:** Software, writing–review and editing. **T. Carver:** Software, writing–review and editing. **A. Lee:** Software, writing–review and editing. **A.P. Cunningham:** Software, writing–review and editing. **M. Lush:** Software, writing–review and editing. **J. Dennis:** Formal analysis, writing–review and editing. **M. Tischkowitz:** Resources, writing–review and editing. **K. Downes:** Resources, writing–review and editing. **D. Hu:** Resources, writing–review and editing. **E. Hahnen:** Resources, writing–review and editing. **R.K. Schmutzler:** Resources, writing–review and editing. **T.L. Stockley:** Resources, writing–review and editing. **G.S. Downs:** Resources, writing–review and editing. **T. Zhang:**

References

1. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med 2019;21:1708–18.
2. Lee A, Mavaddat N, Cunningham A, Carver T, Ficorella L, Archer S, et al. Enhancing the BOADICEA cancer risk prediction model to incorporate new data on RAD51C, RAD51D, BARD1, updates to tumor pathology and cancer incidence. J Medical Genet 2022;59:1206–18.
3. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 2019;104:21–34.
4. Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. Br J Cancer 2008;98:1457–66.
5. Kapoor PM, Mavaddat N, Choudhury PP, Wilcox AN, Lindström S, Behrens S, et al. Combined associations of a polygenic risk score and classical risk factors with breast cancer risk. J Natl Cancer Inst 2021;113:329–37.
6. Rudolph A, Song M, Brook MN, Milne RL, Mavaddat N, Michailidou K 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;47:526–36.
7. Barrdahl M, Canzian F, Joshi AD, Travis RC, Chang-Claude J, Auer PL, et al. Post-GWAS gene-environment interplay in breast cancer: results from the Breast and Prostate Cancer Cohort Consortium and a meta-analysis on 79,000 women. Hum Mol Genet 2014;23:5260–70.
8. Travis RC, Reeves GK, Green J, Bull D, Tipper SJ, Baker K, et al. Gene–environment interactions in 7,610 women with breast cancer: prospective evidence from the Million Women Study. Lancet 2010;375:2143–51.
9. Kapoor PM, Lindström S, Behrens S, Wang X, Michailidou K, Bolla MK, et al. Assessment of interactions between 205 breast cancer susceptibility loci and 13 established risk factors in relation to breast cancer risk, in the Breast Cancer Association Consortium. Int J Epidemiol 2020;49:216–32.
10. Wang X, Chen H, Kapoor PM, Su Y-R, Bolla MK, Dennis J, et al. A genome-wide gene-based gene–environment interaction study of breast cancer in more than 90,000 women. Cancer Research Comm 2022;2:211–9.
11. Brooks JD, Nabi HH, Andrulis IL, Antoniou AC, Chiquette J, Després P, et al. Personalized risk assessment for prevention and early detection of breast cancer: integration and implementation (PERSPECTIVE I&I). J Pers Med 2021;11:511.
12. Archer S, Fennell N, Colvin E, Laquindanum R, Mills M, Dennis R et al. Personalized risk prediction in hereditary breast and ovarian cancer: a protocol for a multicenter randomized controlled trial. Cancers 2022;14:2716.
13. Shieh Y, Eklund M, Madlensky L, Sawyer SD, Thompson CK, Stover Fiscalini A, et al. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. J Natl Cancer Inst 2017;109.
14. Dudbridge F, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, et al. Adjustment for index event bias in genome-wide association studies of subsequent events. Nat Commun 2019;10:1561.
15. Fritsche LG, Ma Y, Zhang D, Salvatore M, Lee S, Zhou X, et al. On cross-ancestry cancer polygenic risk scores. PLoS Genet 2021;17:e1009670.
16. Du Z, Gao G, Adedokun B, Ahearn T, Lunetta KL, Zirpoli G, et al. Evaluating polygenic risk scores for breast cancer in women of African ancestry. J Natl Cancer Inst 2021;113:1168–76.
17. Liu C, Zeinomar N, Chung WK, Kiryluk K, Gharavi AG, Hripcsak G, et al. Generalizability of polygenic risk scores for breast cancer among women with European, African, and Latinx ancestry. JAMA Netw Open 2021;4:e2119084.
18. Ho W-K, Tai M-C, Dennis J, Tai MC, Mariapun S, Li J, et al. Polygenic risk scores for prediction of breast cancer risk in Asian populations. Genet Med 2022;24:586–600.
19. Li SX, Milne RL, Nguyen-Dumont T, Wang X, English DR, Giles GG, et al. Prospective evaluation of the addition of polygenic risk scores to breast cancer risk models. JNCI Cancer Spectr 2021;5:pkab021.
20. Lakeman IMM, Rodríguez-Girondo M, Lee A, Ruiter R, Stricker BH, Wijnant SRA, 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;22:1803–11.
21. Pal Choudhury P, Brook MN, Hurson AN, Lee A, Mulder CV, Coulson P, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. Breast cancer Res 2021;23:22.
22. Hurson AN, Choudhury PP, Gao C, Hüsing A, Eriksson M, Shi M, 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 2022;50:1897–911.
23. Yang X, Eriksson M, Czene K, Lee A, Leslie G, Lush M, et al. Prospective validation of the BOADICEA multifactorial breast cancer risk prediction model in a large prospective cohort study. J Med Genetics 2022;59:1196–205.
24. Carver T, Hartley S, Lee A, Cunningham AP, Archer S, Babb de Villiers C, et al. CanRisk Tool: a web interface for the prediction of breast and ovarian cancer risk and the likelihood of carrying genetic pathogenic variants. Cancer Epidemiol Biomarkers Prev 2021;30:469–73.