Incorporating alternative Polygenic Risk Scores into the BOADICEA breast cancer risk prediction model

Mavaddat N1, Ficorella L1, Carver T1, Lee A1, Cunningham, AP1, Lush M1, Dennis J1, Tischkowitz M2, Downes K3, Donglei H4, Hahnen E5,6, Schmutzler R5-7, Stockley T8-10, Downs G8,10, Zhang, T8, Chiarelli AM11,12, Bojesen S13-15, Cong L16, Chung WK17, Pardo M18, Feliubadaló L19-21, Balmana J18,22, Simard J23, Antoniou AC1, Easton DF1,24

1 Centre for Cancer Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

2 Department of Medical Genetics and National Institute for Health Research, Cambridge Biomedical Research Centre, The University of Cambridge, Cambridge, UK.

3 Cambridge Genomics Laboratory, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK.

4 Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, California, USA.

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.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
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, ON M5S 1A1, Canada.

12 Ontario Health, Cancer Care Ontario, Toronto, ON M5G 2L3, 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, USA.

17 Departments of Pediatrics and Medicine, Columbia University, New York, NY, USA.

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.
Medical Oncology Department, University Hospital of Vall d'Hebron, Barcelona, Spain.

Department of Molecular Medicine, Université Laval and CHU de Québec-Université Laval Research Center

Centre for Cancer Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

Correspondence:

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, UK

Email: nm274@medschl.cam.ac.uk
Phone: +44 (0)1223 748627; Fax: +44 (0)1223 748628
Running title: Incorporating alternative PRS within risk prediction models

Keywords: cancer, breast, risk, prediction, polygenic.

Word count: abstract 250 words, main article: 2280 words

Number of figures and tables: 1 Figure, 1 Table, 12 Supplementary Tables

Funding: This work has been supported by grants from Cancer Research UK (PPRPGM-Nov\textsuperscript{2}0100002); the European Union’s Horizon 2020 research and innovation programme under grant agreement numbers 633784 (B-CAST) and 634935 (BRIDGES); the PERSPECTIVE I&I project which is funded by the Government of Canada through Genome Canada (#13529) and the Canadian Institutes of Health Research (#155865), the Ministère de l’Économie et de l’Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation, the CHU de Quebec Foundation and the Ontario Research Fund; and by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014).

BCAC is funded by the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and the PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l’Économie et de l’Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Additional funding for BCAC is provided via the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health.
Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer Research UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l’Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. MT was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and Cancer Research UK C22770/A31523 (International Alliance for Cancer Early Detection programme).

The PRISMA study was supported by the Carlos III National Health Institute and Ministerio de Educación y Ciencia grant (PI19/01195); and CERCA Programa/Generalitat de Catalunya for institutional support (J. Balmaña).

W. Chung was funded by NIH grant U01 HG008680 (eMERGE grant)

Competing interests: All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare that: The BOADICEA model has been licensed to Cambridge Enterprise for commercialization, with the authors D.F.E., A.C.A., A.P.C., A.L. and T.C. listed as its inventors. A.P. has received royalties as a result of this. A.L. is an employee of Illumina Inc. and owns shares in Illumina Inc. A.A. T.C., J. D. and N.M. are funded by CanRisk CR-UK programme (PPRPGM-Nov20\100002) paid to the institution. N.M. is also funded through the PERSPECTIVE program. S.J. leads the PERSPECTIVE I&I project, which is funded by the Government of Canada through Genome Canada (#13529) and the Canadian Institutes of Health Research (#155865), the Ministère de l’Économie et de l’Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation, the CHU de Quebec Foundation and the Ontario Research Fund (Payments made to institution Université Laval). SJ leads the PERSPECTIVE project.
supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l’Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation (Payments made to institution Université Laval). M.T. is funded by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and Cancer Research UK C22770/A31523 (International Alliance for Cancer Early Detection programme). W.C. is funded by NIH grant U01 HG008680 (eMERGE grant). W.C. received consulting fees from and is on the scientific advisory board of Regeneron Genetics Center and is also on the board of directors of Prime Medicine. S.B. is funded by Region H and Boserup Foundation. These funders did not have any influence on the sample collection, lab work, data analyses or their interpretation. LFeliubadalo declares honoraria for lecture on analysis of sequencing data from ovarian cancer presented to Astra-Zeneca. Other authors report no conflict of interest

Abstract

Background: The multifactorial risk prediction model BOADICEA 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, standard deviation (SD), and proportion of the overall polygenic component explained by the PRS ($\alpha^2$) need to be estimated. $\alpha$ was estimated using logistic regression, where the age-specific log-relative risk 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 $\alpha$, as compared with the RL estimates. The RL $\alpha$ estimates were very close to those obtained by assuming proportionality to the odds ratio 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 enable comprehensive breast cancer risk assessment.
Introduction

BOADICEA\textsuperscript{1,2} is a risk prediction algorithm for predicting breast and ovarian cancer risk on the basis of genetic and non-genetic factors. The algorithm incorporates the effects of common genetic variants, summarised 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)\textsuperscript{1,2} has been specifically developed to incorporate the 313 SNP PRS of Mavaddat et al.\textsuperscript{3}; this PRS was developed using the very large dataset of the Breast Cancer Association Consortium (BCAC) and extensively validated in prospective studies. However, the algorithm itself is flexible and can incorporate any PRS for which the relevant parameters are known. These parameters are the mean ($\mu$) and standard deviation ($\sigma$) of the PRS in the population, and the proportion ($\alpha^2$) of the polygenic variance attributable to the PRS. In practice, the PRS can be normalised and supplied as a Z-score, in which case only parameter $\alpha$ 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 several 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\textsuperscript{1,4}:

$$\lambda(t) = \lambda_0(t) \exp (\delta_{g(i)}(t) + \sigma_p(t) x_p^{(i)} + \sum \beta_p z_p l)$$

(1)
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-hazard ratio associated with genotype $k$). $\sigma_p(t)x_p^{(i)}$ models the polygenic component, $\sigma_p(t)$ being the polygenic standard deviation and $x_p^{(i)}$ the normalised 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 $\gamma$ and $\theta$ 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$^3$:

$$\sigma_K^2(t) = \alpha^2(\gamma + \theta t) \tag{2}$$

$\sigma_K(t)$ can also be interpreted as the age-specific log-hazard ratio per unit SD of the PRS, conditional on other risk factors. Note that in Mavaddat et al.'s equation (1) is written $\sigma_K^2(t) = \gamma^2(\alpha + \beta t)$. The change of symbols is for consistency with Lee et al.$^1$ and the Canrisk platform (www.canrisk.org), where the proportion of the polygenic variance explained by the PRS is denoted as $\alpha^2$.

**Estimation of $\alpha$ and incorporating alternative PRS**

The key parameter is $\alpha$. The first approach to estimating this parameter makes the simplifying assumption that the polygenic standard deviation $\sigma_K(t)$ can be approximated by
the marginal age-specific log-hazard ratio per unit SD of the PRS$^3$. This can then be estimated using case-control (or cohort) data, by first transforming the PRS using:

$$S' = x_k \sqrt{Y + \theta t}$$  (3)

where $x_k$ is the standardised (per unit SD) version of the proposed PRS. $S'$ is then included as a covariate in a logistic regression model: the parameter (log-relative risk) estimate corresponding to the covariate $S'$ gives the required $\alpha$ parameter, which we denote $\alpha_{GLM}$. This can then be estimated using case-control (or cohort) data. 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 et al$^3$ (Supplementary Tables 1 and 2). 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 (PCs).

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 non-genetic risk factors, which have relatively small effects on risk and appear to be independent of the PRS$^5$. However, it may not be true for other genetic factors, in particular the unmeasured polygenic component. To address this problem, we also estimated $\alpha$ using a retrospective likelihood approach ($\alpha_{RL}$), applied to the same BCAC dataset. In this analysis, the observed PRS is computed conditional on the phenotypes of the individuals (age of diagnosis and case/control status). 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).

Since 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 maximised using the optimize function in R. 95% Confidence Intervals were obtained using a grid of values for $\alpha_R$, and finding the difference between the log-likelihoods and the maximum log-likelihood.

Using the value for the proportionality estimated in our data set, we now note that there is a simple proportional relationship between the PRS effect size and alpha. The effect size of a PRS is typically expressed in terms of the log-relative risk per unit SD, say $\eta$. From equations (1) and (2) it can be seen that, under the rare disease assumption, the marginal hazard ratio associated with the PRS should equal the conditional hazard ratio. If differential age effects can also be ignored, $\alpha$ should therefore be approximately proportional to $\eta$. This allows $\alpha$ to be estimated using PRS313 as a standard. Thus: $\alpha_{AP} = \frac{\eta}{\eta_0} \alpha_0$ where $\eta_0$ and $\alpha_0$ are the corresponding estimates for PRS313. This provides a simple method that could be applied to PRS developed and validated on a different dataset.

**PRS Examples**

We computed the relevant parameters for 11 PRS (Supplementary Tables 3-12; SNP positions based on Genome Reference Consortium Human Build 37 (GRCh37)). Six of these are modifications of the PRS313, designed for clinical implementation. The BRIDGES PRS was developed as an 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/](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. 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 randomised 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. 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](https://emerge-network.org)). DBDS299, using data from the Danish Blood Donor Study ([https://bmjopen.bmj.com/content/9/6/e028401](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 et al.$^3$ using Lasso penalised regression, and two PRS (WISDOM75 and WISDOM120) developed for the WISDOM clinical trial$^8$ (Clinical Trials identifier NCT02620852).

The means and standard deviations of each PRS, and the proportion ($\alpha$) of the polygenic variance attributable to these alternative PRS were derived in the same data set, namely the validation and prospective sets described by Mavaddat et al.$^3$
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).

**Ethics Declaration**

All studies 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 available on request.

**Results**

Table 1 summarises 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 $\alpha_{RL}$ parameters 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 $\alpha$) were smaller than for PRS313, while PRS3820 had larger effect sizes. The two WISDOM and PRISMA PRSs also had somewhat smaller effect sizes. Removal of the 2 chromosome 22 SNPs had only a small effect on the estimated log-OR per 1SD, and $\alpha$ – for example reducing
\( \alpha_{RL} \) from 0.441 to 0.439 for PRS313. The \( \alpha \) values computed using the simpler logistic regression approach (\( \alpha_{GLM} \)) were smaller than those generated using the retrospective likelihood approach for all PRS.

Assuming that the \( \alpha \) values are proportional to the log-relative risks per 1 SD, and using the PRS313 as the standard, the predicted value is given by \( \alpha = 0.887 \eta \) (Figure 1). These predicted values were very similar to the \( \alpha_{RL} \) values for all PRS.

**Discussion**

We evaluated approaches to incorporating alternate breast cancer PRSs into the risk prediction algorithm BOADICEA. The \( \alpha \) values computed using the simpler logistic regression approach (\( \alpha_{GLM} \)) were consistently smaller than those generated using the retrospective likelihood approach (\( \alpha_{RL} \)), for all PRS. This is consistent with the fact that women 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, which leads to an underestimation of the PRS effect size if the latter is not allowed for, a phenomenon related to index event bias\(^9\).

We showed further that the \( \alpha \) parameters derived from the log-relative risk estimate by assuming proportionality were very close to the \( \alpha_{RL} \) 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 dataset if a log-relative risk 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 emphasises the importance of correctly normalising 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. 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 emphasises the importance of calibrating the PRS to the relevant population distribution.

The analyses used here adjusted the PRS for both the country in which the study was conducted and ancestry informative principal components. An adjustment is necessary since 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 datasets 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 emphasised that the main validations of BOADICEA utilised PRS313. For PRS that are substantially different, and particularly as more informative PRS are generated through larger GWAS, further validations 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-relative risk 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 PRS31314-17 in addition to other risk factors. It is available through the CanRisk (www.canrisk.org) tool18 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 methods described here allow other PRSs to be used with BOADICEA and hence provide comprehensive breast cancer risk assessment.

References


doi:10.1093/ije/dyab036

Table 1. Summary parameters for alternative PRSs.

Figure 1. Graph of BOADICEA polygenic proportion ($\alpha_{RL}$) vs log OR per 1 SD

Graph of BOADICEA polygenic proportion alpha ($\alpha_{RL}$) vs log OR per 1 SD for PRS313 and alternative PRS. $\alpha_{RL}$ were estimated using the retrospective likelihood method as described in the Methods and Supplementary Methods.

Supplementary Table 1. Studies and samples used in these analyses.

Supplementary Table 2. Country in which the studies were conducted.

Supplementary Table 3. List of SNPs used in each of the PRS.

Supplementary Table 4. SNPs and weights for the BCAC PRS313.

Supplementary Table 5. SNPs and weights for the BRIDGES PRS.

Supplementary Table 6. SNPs and weights for the PERSPECTIVE I&I PRS.

Supplementary Table 7. SNPs and weights for the EASTGLH PRS.

Supplementary Table 8. SNPs and weights for the PRISMA PRS.

Supplementary Table 9. SNPs and weights for the eMERGE PRS.

Supplementary Table 10. SNPs and weights for the DBDS299 PRS.

Supplementary Table 11. SNPs and weights for the WISDOM75 PRS.

Supplementary Table 12. SNPs and weights for the WISDOM128 PRS.
Table 1. Summary parameters for alternative PRSs

<table>
<thead>
<tr>
<th></th>
<th>No of SNPs</th>
<th>Mean controls</th>
<th>Mean cases</th>
<th>SD controls</th>
<th>SD cases</th>
<th>Log OR per 1 SD</th>
<th>95%CI</th>
<th>( \alpha_{GLM} ) a</th>
<th>95%CI</th>
<th>( \alpha_{RL} ) b</th>
<th>95%CI</th>
<th>Ratio (( \alpha_{RL}: \log\ OR\ per\ 1\ SD ))</th>
<th>Predicted ( \alpha_{RL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCAC PRS313</td>
<td>313</td>
<td>-0.424</td>
<td>-0.114</td>
<td>0.611</td>
<td>0.619</td>
<td>0.497</td>
<td>0.476 - 0.519</td>
<td>0.397</td>
<td>0.378 - 0.415</td>
<td>0.441</td>
<td>0.430 - 0.455</td>
<td>0.887</td>
<td>0.441</td>
</tr>
<tr>
<td>BRIDGES</td>
<td>306</td>
<td>-0.422</td>
<td>-0.114</td>
<td>0.608</td>
<td>0.615</td>
<td>0.495</td>
<td>0.474 - 0.517</td>
<td>0.394</td>
<td>0.376 - 0.413</td>
<td>0.439</td>
<td>0.425 - 0.450</td>
<td>0.886</td>
<td>0.439</td>
</tr>
<tr>
<td>PERSPECTIVE</td>
<td>295</td>
<td>-0.448</td>
<td>-0.147</td>
<td>0.599</td>
<td>0.609</td>
<td>0.489</td>
<td>0.468 - 0.511</td>
<td>0.389</td>
<td>0.371 - 0.408</td>
<td>0.433</td>
<td>0.420 - 0.450</td>
<td>0.886</td>
<td>0.434</td>
</tr>
<tr>
<td>EASTGLH</td>
<td>303</td>
<td>-0.407</td>
<td>-0.100</td>
<td>0.606</td>
<td>0.613</td>
<td>0.494</td>
<td>0.472 - 0.516</td>
<td>0.394</td>
<td>0.375 - 0.412</td>
<td>0.438</td>
<td>0.420 - 0.450</td>
<td>0.886</td>
<td>0.438</td>
</tr>
<tr>
<td>PRISMA</td>
<td>268</td>
<td>0.322</td>
<td>0.575</td>
<td>0.558</td>
<td>0.565</td>
<td>0.446</td>
<td>0.424 - 0.467</td>
<td>0.358</td>
<td>0.339 - 0.376</td>
<td>0.396</td>
<td>0.385 - 0.410</td>
<td>0.890</td>
<td>0.395</td>
</tr>
<tr>
<td>eMERGE</td>
<td>308</td>
<td>-0.456</td>
<td>-0.150</td>
<td>0.608</td>
<td>0.616</td>
<td>0.495</td>
<td>0.473 - 0.516</td>
<td>0.394</td>
<td>0.376 - 0.413</td>
<td>0.439</td>
<td>0.425 - 0.450</td>
<td>0.887</td>
<td>0.439</td>
</tr>
<tr>
<td>DBDS299</td>
<td>299</td>
<td>-0.508</td>
<td>-0.211</td>
<td>0.596</td>
<td>0.605</td>
<td>0.486</td>
<td>0.465 - 0.508</td>
<td>0.387</td>
<td>0.369 - 0.406</td>
<td>0.432</td>
<td>0.440 - 0.455</td>
<td>0.888</td>
<td>0.431</td>
</tr>
<tr>
<td>BCAC PRS77</td>
<td>77</td>
<td>-0.892</td>
<td>-0.703</td>
<td>0.449</td>
<td>0.460</td>
<td>0.394</td>
<td>0.373 - 0.415</td>
<td>0.310</td>
<td>0.292 - 0.328</td>
<td>0.350</td>
<td>0.340 - 0.360</td>
<td>0.889</td>
<td>0.349</td>
</tr>
<tr>
<td>BCAC PRS3820</td>
<td>3820</td>
<td>-0.445</td>
<td>-0.199</td>
<td>0.460</td>
<td>0.463</td>
<td>0.518</td>
<td>0.496 - 0.540</td>
<td>0.412</td>
<td>0.393 - 0.431</td>
<td>0.454</td>
<td>0.440 - 0.470</td>
<td>0.878</td>
<td>0.459</td>
</tr>
<tr>
<td>WISDOM-75</td>
<td>74</td>
<td>-1.057</td>
<td>-0.835</td>
<td>0.567</td>
<td>0.573</td>
<td>0.360</td>
<td>0.338 - 0.381</td>
<td>0.281</td>
<td>0.263 - 0.299</td>
<td>0.315</td>
<td>0.300 - 0.330</td>
<td>0.875</td>
<td>0.319</td>
</tr>
<tr>
<td>WISDOM-128</td>
<td>126</td>
<td>-0.180</td>
<td>0.041</td>
<td>0.464</td>
<td>0.478</td>
<td>0.449</td>
<td>0.427 - 0.470</td>
<td>0.355</td>
<td>0.337 - 0.374</td>
<td>0.402</td>
<td>0.390 - 0.415</td>
<td>0.895</td>
<td>0.398</td>
</tr>
</tbody>
</table>

Excluding SNPs correlated with CHEK2*1100delC

<table>
<thead>
<tr>
<th></th>
<th>No of SNPs</th>
<th>Mean controls</th>
<th>Mean cases</th>
<th>SD controls</th>
<th>SD cases</th>
<th>Log OR per 1 SD</th>
<th>95%CI</th>
<th>( \alpha_{GLM} ) a</th>
<th>95%CI</th>
<th>( \alpha_{RL} ) b</th>
<th>95%CI</th>
<th>Ratio (( \alpha_{RL}: \log\ OR\ per\ 1\ SD ))</th>
<th>Predicted ( \alpha_{RL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCAC PRS311</td>
<td>311</td>
<td>-0.092</td>
<td>0.217</td>
<td>0.609</td>
<td>0.618</td>
<td>0.496</td>
<td>0.474 - 0.518</td>
<td>0.395</td>
<td>0.377 - 0.414</td>
<td>0.439</td>
<td>0.425 - 0.455</td>
<td>0.886</td>
<td>0.440</td>
</tr>
<tr>
<td>BRIDGES</td>
<td>305</td>
<td>-0.428</td>
<td>-0.121</td>
<td>0.607</td>
<td>0.615</td>
<td>0.495</td>
<td>0.473 - 0.516</td>
<td>0.394</td>
<td>0.375 - 0.412</td>
<td>0.438</td>
<td>0.425 - 0.450</td>
<td>0.886</td>
<td>0.439</td>
</tr>
<tr>
<td>PERSPECTIVE</td>
<td>293</td>
<td>-0.116</td>
<td>0.183</td>
<td>0.598</td>
<td>0.607</td>
<td>0.488</td>
<td>0.466 - 0.510</td>
<td>0.388</td>
<td>0.369 - 0.406</td>
<td>0.432</td>
<td>0.420 - 0.445</td>
<td>0.885</td>
<td>0.433</td>
</tr>
<tr>
<td>EASTGLH</td>
<td>301</td>
<td>-0.075</td>
<td>0.230</td>
<td>0.604</td>
<td>0.612</td>
<td>0.492</td>
<td>0.471 - 0.514</td>
<td>0.392</td>
<td>0.374 - 0.411</td>
<td>0.436</td>
<td>0.420 - 0.450</td>
<td>0.885</td>
<td>0.437</td>
</tr>
<tr>
<td>eMERGE</td>
<td>306</td>
<td>-0.124</td>
<td>0.181</td>
<td>0.606</td>
<td>0.614</td>
<td>0.493</td>
<td>0.471 - 0.515</td>
<td>0.393</td>
<td>0.374 - 0.411</td>
<td>0.437</td>
<td>0.425 - 0.450</td>
<td>0.886</td>
<td>0.437</td>
</tr>
<tr>
<td>DBDS299</td>
<td>297</td>
<td>-0.177</td>
<td>0.119</td>
<td>0.594</td>
<td>0.603</td>
<td>0.485</td>
<td>0.463 - 0.506</td>
<td>0.386</td>
<td>0.367 - 0.404</td>
<td>0.430</td>
<td>0.387</td>
<td>0.378</td>
<td>0.430</td>
</tr>
<tr>
<td>BCAC PRS3818</td>
<td>3818</td>
<td>-0.227</td>
<td>0.018</td>
<td>0.459</td>
<td>0.462</td>
<td>0.517</td>
<td>0.495 - 0.539</td>
<td>0.411</td>
<td>0.392 - 0.430</td>
<td>0.453</td>
<td>0.440 - 0.465</td>
<td>0.877</td>
<td>0.458</td>
</tr>
</tbody>
</table>

Results based on 22767 controls and 16151 cases in the prospective and validation sets from Mavaddat et al., AJHG (2020) European women with known age and age less than 80; SD, standard deviation; CI, Confidence Intervals.

a\( \alpha_{GLM} \) is based on logistic regression adjusted for country in which studies were conducted and 10 principal components.

b\( \alpha_{RL} \) is based on retrospective likelihood method as described in the 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.