Enhancing the BOADICEA cancer risk prediction model to incorporate new data on *RAD51C, RAD51D, BARD1,* updates to tumour pathology and cancer incidence.

Andrew Lee[1], Nasim Mavaddat[1], Alex P. Cunningham[1], Tim Carver[1], Lorenzo Ficorella[1], Stephanie Archer[2], Fiona M. Walter[2,3], Marc Tischkowitz[4], Jonathan Roberts[4], Juliet Usher-Smith[2], Jacques Simard[5], Marjanka K. Schmidt[6], Peter Devilee[7], Vesna Zadnik[8], Hannes Jürgens[9], Emmanuelle Mouret-Fourme[10], Antoine De Pauw[11], Matti Rookus[12], Thea M. Mooij[12], Paul P.D. Pharoah[1,13], Douglas F. Easton[1,13], Antonis C. Antoniou[1,†]

[1]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, The University of Cambridge, Strangeways Research Laboratory, Cambridge CB1 8RN, UK

[2]The Primary Care Unit, Department of Public Health & Primary Care, The University of Cambridge, Cambridge, CB1 8RN, UK

[3]Wolfson Institute of Population Health, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, UK

[4]Department of Medical Genetics and National Institute for Health Research, Cambridge Biomedical Research Centre, The University of Cambridge, Cambridge, UK

[5]Centre Hospitalier Universitaire de Québec–Université Laval Research Center, 2705 Laurier Boulevard, Québec City, QC, G1V4G2, Canada

[6]Division of Molecular Pathology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands

[7]Department of Human Genetics, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands

[8]Epidemiology and Cancer Registry, Institute of Oncology, Ljubljana, Slovenia

[9]Tartu University Hospital, Clinic of Hematology & Oncology, Puusepa 8, Tartu, Estonia

[10]Service de Génétique, Institut Curie, 35, rue Dailly, Saint-Cloud, 92210, France

[11]Service de Génétique, Institut Curie, 26, rue d'Ulm, Paris, 75005, France

[12]Department of Epidemiology, Netherlands Cancer Institute, P.O. Box 90203, 1006 BE Amsterdam, The Netherlands

[13]Centre for Cancer Genetic Epidemiology, Department of Oncology, The University of Cambridge, Cambridge CB1 8RN, UK

[†]**Corresponding author**: Antonis C. Antoniou, Centre for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, CB1 8RN, UK, aca20@medschl.cam.ac.uk, +44 1223 748630
ORCID ID: 0000-0001-9223-3116

Key Words: breast cancer, ovarian cancer, risk, *RAD51C, RAD51D, BARD1, PALB2,* BOADICEA, CanRisk

Word Count: 4102

Abstract

Background

BOADICEA for breast cancer and the epithelial ovarian cancer (EOC) models included in the CanRisk tool (www.canrisk.org) provide future cancer risks based on pathogenic variants in cancer-susceptibility genes, polygenic risk scores, breast-density, questionnaire-based risk factors and family history. Here, we extend the models to include the effects of pathogenic variants in recently established breast cancer and EOC susceptibility genes, up-to-date age-specific pathology distributions and continuous risk factors.

Methods

BOADICEA was extended to further incorporate the associations of pathogenic variants in *BARD1*, *RAD51C* and *RAD51D* with breast cancer risk. The EOC model was extended to include the association of *PALB2* pathogenic variants with EOC risk. Age-specific distributions of oestrogen-receptor-negative and triple-negative breast cancer status for pathogenic variant carriers in these genes and *CHEK2* and *ATM* were also incorporated. A novel method to include continuous risk factors was developed, exemplified by including adult-height as continuous.

Results

BARD1, *RAD51C* and *RAD51D* explain 0.31% of the breast cancer polygenic variance. When incorporated into the multifactorial model, 34-44% of these carriers would be reclassified to the near-population and 15-22% to the high-risk categories based on the UK NICE guidelines. Under the EOC multifactorial model, 62%, 35% and 3% of *PALB2* carriers have lifetime EOC risks of <5%, 5-10% and >10% respectively. Including height as continuous, increased the BC relative-risk variance from 0.002 to 0.010.

Conclusions

These extensions will allow for better personalised risks for *BARD1*, *RAD51C*, *RAD51D* and *PALB2* pathogenic variant carriers and more informed choices on screening, prevention, risk factor modification or other risk-reducing options.

Introduction

Breast cancer (BC) and epithelial tubo-ovarian cancer (EOC) are two of the most common cancers in females[1 2]. Through mammography or other methods, screening for BC can reduce mortality, and organised screening is available in most developed countries[3]. For EOC, no effective screening exists, but the disease can be prevented by salpingo-oophorectomy. However, these preventative options are associated with adverse effects. Therefore, identifying those at increased risk may help to target screening and preventative options to those most likely to benefit[4]. Both BC and EOC risks are multifactorial diseases, with family history of cancer (FH), genetic factors and lifestyle, hormonal and reproductive risk factors (RF) all contributing to risk[5-7].

Previously we developed the BOADICEA (Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm) model for BC risk prediction and for the likelihood of carrying pathogenic variants (PVs) in BC susceptibility genes. BOADICEA v5 incorporates the effects of PVs in five BC susceptibility genes (*BRCA1*, *BRCA2*, *PALB2*, *CHEK2* and *ATM*), the effects of known common genetic variants summarised as a polygenic risk score (PRS, accounting for ~20% of the polygenic variance), and a polygenic component that accounts for any residual familial aggregation[8 9]. We also developed a similar EOC model (Ovarian Cancer Model v1) that considers the effects of PVs in *BRCA1*, *BRCA2*, *RAD51D*, *RAD51C* and *BRIP1* on EOC together with a PRS (accounting for ~5% of the polygenic variance) and a residual polygenic component[10 11]. BOADICEA includes mammographic density and both models incorporate the effects of known lifestyle, hormonal, reproductive and anthropometric RFs. In addition, the models incorporate breast tumour heterogeneity by considering the distributions of tumour oestrogen receptor (ER) and triple-negative (TN) (ER, progesterone receptor and human epidermal growth factor receptor 2 negative) status for *BRCA1* and *BRCA2* PV carriers and the general population[12 13]. Both models are freely available to healthcare professionals via the CanRisk webtool (www.canrisk.org), and are widely used by healthcare professionals[14].

Recently, large population-based and family-based targeted sequencing studies have established that PVs in *RAD51C, RAD51D* and *BARD1* are associated with BC risk[15 16] and that

PVs in *PALB2* are associated with EOC risk[17 18]. In addition, analysis of the tumour characteristics in the BRIDGES study has provided age-specific estimates of the distributions of tumour characteristics for PV carriers in all established susceptibility genes[19].

A further limitation of the previous models is that all epidemiological RFs are treated as categorical. However, some RFs (e.g., height, body mass index (BMI) mammographic density) are intrinsically continuous, and discretisation results in a loss of information, reducing their predictive ability.

Here we extend both models to explicitly model the effects of PVs in the recently established BC and EOC susceptibility genes and incorporate up-to-date age-specific pathology distributions. We present a methodological framework for incorporating continuous RFs into the model, and we demonstrate this by including height as a continuous variable. Finally, we describe updates to the population reference cancer incidence rates used in the models by incorporating more up-to-date incidences, incidences for additional countries and refining the derivation of birth-cohort specific incidences for inclusion in the models that address sparsity in the population incidence data.

Methods

Rare Moderate-Risk Pathogenic Variants

Both BOADICEA and the EOC Model model cancer incidence as an explicit function of PVs in known high- and moderate-penetrance susceptibility genes (major genes) together with a polygenic component[9-12 20-22]. By using an explicit genetic model, they can account for both genetic testing and detailed FH. BOADICEA includes the genes *BRCA1*, *BRCA2*, *PALB2*, *CHEK2* and *ATM*, with dominance in that order, along with a BC susceptibility polygenic component. The EOC Model includes the genes *BRCA1*, *BRCA2*, *RAD51D*, *RAD51C* and *BRIP1*, with dominance in that order, along with an EOC susceptibility polygenic component. Details of the underlying model are included in the supplementary material. The values of the parameters for the original models were determined by complex segregation analysis[9 10]. However, this was not possible for the extended versions since no sufficiently large data set containing all the model features was available. Instead, we adopted a synthetic approach[23], in which additional model parameters are taken from large-scale external studies[8 11 12 21].

Here, BOADICEA was extended to explicitly model the effects of PVs in *BARD1*, *RAD51C* and *RAD51D* (eight BC susceptibility genes), while the EOC model was extended to include *PALB2* (six EOC susceptibility genes). In both models, the effects of PVs were included as major genes and are parameterised by their allele frequency in the general population and their age-specific relative risks (RR). The BC RR for carriers of PV in *BARD1* was taken from the BRIDGES study[15], while those for *RAD51C* and *RAD51D* were the meta-analysed values from Dorling et al[15] and Yang et al[24]. The EOC RR for *PALB2* PV carriers was taken from Yang et al[17]. The BRIDGES study[15] suggested that the RR estimates associated with PVs in *ATM* are lower than the previously assumed estimate of 2.8[21] and it was therefore updated to the Dorling et al[15] estimate. The previously assumed RR estimates for PVs in *BRCA1, BRCA2, PALB2* and *CHEK2*[8 21] were based on large studies that enabled the estimation of age-specific risks, or were estimated as part of the BOADICEA model fitting process, and were not updated, except for the *BRCA2* associated EOC RRs for ages 59 and over (supplementary material). The PV carrier frequencies for *PALB2*, *CHEK2* (including all PVs), *ATM, BARD1, RAD51D*, *RAD51C* and *BRIP1* and screening test sensitivities for all genes were derived from Dorling et al[15]. We used the BRIDGES study to derive these frequency estimates as it is a very large population-based dataset that includes targeted sequencing data. Frequencies were based on the control frequencies in European populations, adjusted for the assumed sensitivity of the sequencing and the fact that large rearrangements were not detectable (supplementary material). The default sensitivities were then calculated, assuming that clinical genetic testing will detect all known pathogenic mutations except for large rearrangements (except *BRCA1* and *BRCA2*, where testing for large rearrangements is routinely done). All model parameters for PVs are given in Table 1.

As the polygenic component captures all residual familial aggregation not explained by the major genes, the previous models implicitly included the contributions of PVs in the new genes (i.e., *BARD1, RAD51C* and *RAD51D* for BOADICEA and *PALB2* for the EOC Model). Therefore, to avoid double counting their contribution, it was necessary to remove their contribution from the polygenic component by adjusting the log-RR per standard deviation of the polygenic component such that the total variance of the polygenic component and the

new genes is the same as that of the polygenic component of the previous model[21] (supplementary material).

The association between *PALB2* PVs and EOC was also included in the BOADICEA model, and the associations with male BC and pancreatic cancer have been included in both models[17].

The effects of PV in the new BC susceptibility genes on risk prediction were assessed by considering the risk categories described in the National Institute for Health and Care Excellence (NICE) familial BC guidelines[25] for hypothetical women with different ages or family history. For lifetime risk (age 20 to 80 years), three categories are defined: 1) near-population-risk, for risks less than 17%, 2) moderate-risk, for risks in the range [17%, 30%) and 3) high-risk, for risks of 30% or greater. Reclassification was considered based on questionnaire-based RFs (QRF) (RFs other than mammographic density), mammographic density (MD, based on the BI-RADS system) and a polygenic risk score (PRS). The assumed distributions and RRs for QRFs and MD have been described in detail previously[8 11]. For BC, the PRS was taken to be the Breast Cancer Association Consortium 313 variant PRS, which accounts for 20% of the overall polygenic variance[8 26]. For EOC, we defined three risk categories based on lifetime risk[27 28]: 1) near-population-risk, for risks of less than 5%, 2) moderate-risk, for risks in the range [5%, 10%) and 3) high-risk, for risks of 10% or greater, and reclassification was considered based on RFs and a PRS. For EOC, the PRS was taken as the Ovarian Cancer Association Consortium 36 variant PRS, which accounts for 5% of the overall polygenic variance[11 29].

Updates to Tumour Pathology

Both models incorporate data on BC tumour pathology, specifically ER and TN. The distribution of pathology for affected carriers of PVs differs substantially from that in non-carriers for several genes, so that pathology data can affect the carrier probabilities and hence cancer risks[11 12]. In BOADICEA and the EOC model, breast tumours are classified into five groups based on ER and TN status: ER unknown, ER-positive, ER-negative/TN unknown, ER-negative/not TN, and TN. Previously, the models achieved this using age-dependent distributions in the general population and *BRCA1* and *BRCA2* PV carriers and an age-

independent distribution for *CHEK2* PV carriers[12 21]. Due to a lack of data, the tumour ER distribution for carriers of PV in other genes was assumed to be the same as the general population. Here, the models have been updated to incorporate age-dependent ER and TN tumour distributions for carriers of PVs in the BC susceptibility genes *PALB2*, *CHEK2*, *ATM*, *BARD1*, *RAD51C* and *RAD51D*, using data from BRIDGES[19].

Continuous Risk Factors

The previous versions of the models included reproductive, lifestyle, hormonal and anthropometric RFs[8 11]. One limitation of these models was that the RFs needed to be coded as categorical variables. Some RFs are naturally continuous, requiring prior discretisation to a finite number of categories, resulting in some loss of information and reduction in risk discrimination. Here, the methodology was extended to allow the inclusion of continuous risk factors.

The key challenge is to calculate the baseline incidences $\lambda_0(t)$ in equation (1) (supplementary material) from the population incidence and the risk factor distributions. The baseline incidences are calculated sequentially for each age t (considered discrete) using the values at age $t-1$, starting from age 0, requiring the evolution with age of the probability distribution of those who are disease-free [30]. For discrete factors/genes, this involves summing over all possible categories/genotypes, but for continuous factors/genes, it would involve integrating over all possible values. In principle, these integrals could be computed (either analytically or numerically). However, at each age, the number of terms in the integrand increases by a factor of 2, so by age 80, there are $>10^{24}$ terms, with evaluation becomes impracticable. Alternatively, the risk factor could be discretised into a very large number of categories. This would give a very close approximation to the continuous distribution, but (particularly once multiple risk factors are considered, as here) the large number of categories would also make the calculations impractical. Instead, we propose an alternate approach in which the continuous factors are discretised with categories adapted according to the observed RF. The approach is as follows:

1. Firstly, discretise the range of possible risk factor values into a finite number (n) of bins and calculate the probability mass and RR for each bin from the probability density and RR function for the continuous RF. This part is identical to the standard approach for discretising risk factors, used in the existing models[8]. For a risk factor, x, with probability density $P(x)$ and relative risk $RR(x)$ the probability mass for bin i with range $[l_i, u_i]$ is:

$$P(i) = \int_{l_i}^{u_i} P(x)\, dx, \tag{1}$$

and the corresponding RR is

$$RR(i) = \frac{1}{P(i)} \int_{l_i}^{u_i} RR(x)\, P(x)\, dx. \tag{2}$$

2. Create an additional $(n+1)^{th}$ bin based on the individual's measured risk factor value that has an infinitesimal width. The RR for this bin is taken as the RR at the measured value, and it has zero mass. As this bin is infinitesimal, its overlap with the other bins is zero, so there is no double-counting.

This procedure creates a categorical risk factor with $n+1$ categories, where the individual is assigned to the $(n+1)^{th}$ category defined in step 2. This allows the exact value of the risk for the individual to be used, while the number of categories required to compute the baseline rates is fixed, limiting the computation time.

The accuracy of the approximation in the procedure relies on the assumption that the range of values within each bin have similar RRs, which should be reflected in the choice of discretisation scheme and the number of bins n. These choices will depend on the shape of the distribution and the RR function.

The above procedure can be applied to any risk factor distribution or RR function. However, the process assumes that an individual's position within the distribution is fixed with respect to age, although the value of the risk factor and RR may vary with age. Here, the method was applied to height.

Updates to Population Incidences

The baseline incidences in equation (1) are birth-year and country specific as a consequence of using birth-year and country-specific population incidences in the constraining process. We refined the derivation of cohort-specific population incidences to account for variability in the incidences due to small numbers. In addition, we have updated existing incidences in the model to include more recent calendar periods and adapted the model to utilise cancer incidence from four new populations: the Netherlands, France, Slovenia and Estonia. Details are included in the supplementary material.

Results

Rare Moderate-Risk Pathogenic Variants

Table 1 summarises the models' genetic parameter estimates, including those for the new genes. The estimated cumulative age-specific BC risks for *BARD1, RAD51C* and *RAD51D* PV carriers in BOADICEA and EOC risks for *PALB2* carriers, assuming the UK incidences applicable to those born in the 1980s, are shown in figure 1. The estimated average lifetime BC risks for PV carriers are 24%, 22% and 21% for *BARD1, RAD51C* and *RAD51D* PV carriers, respectively. The estimated lifetime EOC risk for *PALB2* carriers is 5.0%. Based on the assumed allele frequencies, 0.22% of the population carry PV in the genes *BARD1*, *RAD51C* or *RAD51D,* and these explain on average 0.31% of the female BC polygenic variance (averaged over all ages and cohorts, weighted by the age- and cohort-specific BC incidences). Approximately 0.13% of the population carry PVs in *PALB2,* explaining 0.16% of the EOC polygenic variance and 2.5% of the male BC polygenic variance.

Figure 2 (a)-(f) and Supplementary Table s1 show the distributions of lifetime BC risks for carriers of PVs in B*ARD1*, *RAD51C* and *RAD51D* for a female with unknown FH and a female whose mother is affected at age 50 based on PV carrier status alone and including QRF, MD and a PRS. Based solely on PV carrier status, all females with unknown FH would be classified as at moderate risk. When information on QRF, MD or PRS is known, there is significant reclassification to near-population and high-risk categories, which is greatest when all factors are used in combination. For example, based on lifetime BC risks and using the full

multifactorial model incorporating QRF, MD and a PRS, 33.9% of *BARD1* PV carriers with unknown FH would be reclassified from moderate-risk to near-population-risk, and 21.9% would be reclassified to high-risk (Supplementary Table s1). Similarly, *BARD1* PV carriers with an affected first-degree relative would be considered high-risk (risk of 33.7% by age 80) based on family history and PV status alone. Incorporating the other risk factors would reclassify 12% as near-population risk and 40.2% as moderate-risk (Supplementary Table s1).

Figure 2 (g) and (h) and Supplementary Table s2 show the distribution of lifetime EOC risks for carriers of PVs in *PALB2* for a female with unknown FH and a female whose mother is affected at age 50, as a function of the RFs and PRS. For a *PALB2* carrier with unknown FH, when the RFs and PRS are considered jointly, 62.4% are classified as near-population-risk, 34.9% as moderate-risk, and 2.7% as high-risk. The corresponding proportions with an affected mother are 11.2%, 55.8%, and 33%, respectively. However, even among *PALB2* carriers with an affected mother, 97.5% will have risks of less than 3% by age 50 (Supplementary Table s2).

Tumour Pathology

Figure 3 and Supplementary Tables s3 and s4 show the age-specific distributions of ER-negative tumours and TN tumours among ER-negative tumours used in the models for *PALB2, ATM, CHEK2, BARD1, RAD51C* and *RAD51D* PV carriers based on the BRIDGES data[19]. *BARD1, RAD51C* and *RAD51D PV* carriers predominantly develop ER-negative BCs, and the proportions decrease with increasing age. On the other hand, *CHEK2* and *ATM* carriers primarily develop ER-positive BCs, and the proportion of ER-positive tumours increases with age. Among those with ER-negative tumours, most tumours are TN for PV carriers in all genes, except *CHEK2* carriers, in whom the majority are ER-negative but not TN.

Using the updated age- and gene-specific ER-negative and TN tumour status distributions resulted in differences in the predicted overall and gene-specific carrier probabilities by different tumour pathology and age (Figure 4). For *ATM*, the carrier probabilities for ER-negative tumours are reduced relative to previous estimates, reflecting the stronger association with ER-positive disease. Carrier probabilities for *CHEK2* now show a decline with

age for ER-negative tumours (previously, this was only predicted for ER-positive disease). The carrier probabilities for *PALB2* remain similar to previous estimates. For the new genes *BARD1*, *RAD51C* and *RAD51D*, the carrier probabilities are, as expected, higher for ER-negative and TN disease, but there is little variation by age.

Continuous Risk Factors

As previously, adult female height was assumed to be normally distributed with mean 162.81cm and standard deviation 6.452cm, and be associated a log-RR per standard deviation, for both breast and ovarian cancer, of 0.10130[8 11]. We therefore discretised the normal distribution such that the probability masses of the bins were given by a binomial distribution $B(n-1,\frac{1}{2})$, giving sufficient discretisation to adequately capture the tails of the distribution. We examined the relative discretisation error of the predicted lifetime risk as a function of the number of bins (Figure 5 e and f) and chose $n=5$, as the lowest number of bins such that the root-mean-square relative error was less than 10^{-4}. Compared to the discrete (five-level) RF, the variance of the RR of both BC and EOC increased from 0.002 to 0.010 when height was included as a continuous RF. The effects on predicted lifetime risks are shown in Figure 5 (a)-(d). Under the continuous implementation here, the lifetime BC risk varied from 9.7% for the first percentile to 14.6% for the 99th, whereas under the previous discrete distribution, the risks range from 10.1% to 14.2%.

Discussion

This work has extended the multifactorial BOADICEA BC and EOC risk prediction models (BOADICEA v6 and the Ovarian Cancer Model v2), employing a synthetic approach[23]. The explicit effects of PVs in *RAD51C, RAD51D, BARD1* and *PALB2,* which have now been established as BC and/or EOC susceptibility genes[15 16 17] and are commonly included on cancer gene panels, are now included in the models. The models have also been extended to accommodate continuous RFs, and parameterisation of tumour pathology and cancer incidence have been updated with more recent data. These represent the most comprehensive models for BC and EOC and will allow more complete BC and EOC risk assessment of those undergoing gene-panel testing. In a separate study, the BOADICEA v6

breast cancer model presented here, has been validated in an independent prospective study of 66,415 women attending mammographic screening in Sweden. The full model, including RFs, mammographic density, PRS and PVs in *BRCA1, BRCA2, PALB2, CHEK2, ATM, BARD1, RAD51C* and *RAD51D* was well calibrated overall (calibration slope 0.97 (95%CI:0.95-0.99) and in deciles of predicted 5-year risks; and had a C-index of 0.71(95%CI:0.68-0.74) for discriminating between affected and unaffected women [Yang et al, under review].

By explicitly modelling the effects of PVs in the new cancer susceptibility genes, the models provide personalised cancer risks of PV carriers when combined with QRFs, MD and PRS. Although the numbers affected by these changes will be small at population level, for individuals with *RAD51C, RAD51D* an *BARD1* PVs and their families, the updated risks will be clinically important. *RAD51C*, *RAD51D* and *BARD1* (like *ATM* and *CHEK2*) would be classified as "moderate risk" BC genes based on the average risks[15]. However, according to the BOADICEA predictions, over half (56-59%) of carriers of PVs in these genes in the population would be reclassified from moderate BC risk category to either near-population-risk (34-44%) or high-risk (15-22%), if data on the other risk factors were incorporated (Table s1). Such changes may have important implications for discussions around earlier or more frequent screening or on risk-reduction options for these women. Similarly, based on the multifactorial EOC model ~38% of *PALB2* PV carriers will have lifetime EOC risks of >5% (Table s2), which may influence recommendations on the timing of risk-reducing surgery.

As previously, the models assume that the effects of the PVs in the new genes interact multiplicatively with the PRS and the RFs. No studies have yet assessed the joint effects of PVs in these genes and the PRS or RFs. Previous results for *CHEK2* and *ATM* suggest that the multiplicative model holds true for earlier versions of the PRS[31-33]. Unlike *CHEK2* and *ATM*, however, the new genes predispose more strongly to ER-negative disease, and the combined effect may depart from the multiplicative assumption. Demonstrating this explicitly for the new genes will be challenging given the rarity of the mutations. The multiplicative model has also been shown to be reasonable for the combined effects of PRS and RFs[34], but there is as yet no large-scale evaluation of the combined effects of PVs and RFs. However, recent prospective validation studies of the current and previous versions of the models suggest that, overall, the models fit well[11][Yang et al. under review]. Should deviations from the

multiplicative model between these PVs and RFs emerge, the model can be updated to take them into account.

Both the BC and EOC models incorporate PVs' effects using the estimated population allele frequencies and RRs. These are combined with reference population incidences to calculate absolute risks while constraining the overall incidences over the RFs included in the model. Our implementation used RR and allele frequency estimates from the largest available studies on those of European ancestry[15]. These were assumed to be constant across all countries. Available data are currently too sparse to obtain country-specific estimates. Although there is no evidence that RRs vary among populations, the allele frequencies are likely to vary to some extent [15]. This is most apparent for *CHEK2*, where the founder c.1100delC variant (p.Thr367Metfs*15) is common in northwest Europe with carrier frequencies between 0.3% to 1.2%[35] and explains the majority of carriers but is rare or absent in other populations. If population-specific variant frequencies can be generated, the model can be easily updated to accommodate these. Nevertheless, by allowing population incidences to vary by country, the predicted absolute risks given by the models are country-specific.

The updated age-specific distributions of tumour ER and TN status for six of the BC susceptibility genes in the model (*PALB2*, *CHEK2*, *ATM*, *BARD1*, *RAD51C* and *RAD51D*) should allow better differentiation between PVs that may be present in a family and provide age- and gene-specific mutation carrier probabilities consistent with the prevalence of PVs observed in Mavaddat et al[19]. We note however that estimates are more uncertain at very young and very old ages, where the data are sparse, and more extensive validation may be required in these age-groups. Since PV carrier probabilities are used internally in the models, these will also impact the predicted absolute risks for all unaffected individuals if information on tumour characteristics is available for affected relatives whether or not they carry a PV.

We have developed a novel methodological approach for including continuous RFs into the models. We demonstrated this by including height in both the BC and EOC models, allowing for more nuanced predictions and improving the risk discrimination. While the resulting discrimination based on height alone is modest, the framework will allow other more predictive RFs to be included in the model if accurate risk estimates become available. The

most important example is MD: continuous measures of MD, available through tools such as STRATUS, CUMULUS and Volpara[36-38], have been shown to have stronger associations with BC risk than the categorical BI-RADS system. Other examples include BMI and ages at menarche and menopause. Further, the method could be applied to the joint distribution of several continuous risk factors, where the integrals in equations (1) and (2) become multidimensional integrals.

We have further refined the method for creating cohort incidences from calendar period incidences (supplementary material). The approach provides incidences that are less sensitive to year-on-year fluctuations by averaging over all years in the birth cohort. This method is particularly useful for cancers with low incidences, such as EOC and male BC, where the population size is small, and there is no prior averaging over calendar years. The refinement will have little effect on incidences from larger countries.

Our models have certain limitations. No single dataset containing all the required information was available to construct the multifactorial models, so the models were extended via a synthetic approach. The new model parameters were taken from extensive, well-designed published studies together with existing parameters from model fitting[9 10]. We and others have used this approach for developing previous versions of the models[8 11 12 21 39 40], which have been shown to provide clinically valid predictions[41 42][Yang et al, under review]. As is the case for the previous versions, the updates presented here are primarily based on studies of those of European ancestry in developed countries. There is little evidence that the RRs associated with PVs differ by ancestry. The PV frequencies are also broadly similar across populations, except for specific founder mutations and *CHEK2* PVs, which have a much higher frequency in European than non-European populations. However, other parameters in the model, including RF and PRS distributions, will differ by population, and the model will need to be adapted for use in non-European ancestry populations and developing countries. The synthetic approach presented here allows the model to be easily customised to other populations as better estimates become available[43 44]. Although we used the associations between PVs and tumour ER and TN status, the models do not currently consider the associations with intrinsic BC subtypes based on combinations of ER, progesterone receptor, HER2 and/or grade[19]. The methodology described here could be used to further extend the

models to consider these BC subtypes. Finally, the models make the simplifying assumption that PVs in the assumed BC and EOC susceptibility genes are associated with similar risks to those for truncating variants. These would include missense variants which have similar risks to truncating variants. However, there is evidence that missense variants in *CHEK2* and *ATM* are associated with BC risk, which may be different from the risks for truncating variants[45]. The models would not be applicable to carriers of such variants.

The new model features have been built on the established and well-validated BOADICEA and EOC models[8 11 41]. The updated models will allow for more personalised risk assessment and can help guide decisions on screening, prevention, risk factor modification or other risk-reducing options. The models presented are now available for use by healthcare professionals through the user-friendly CanRisk webtool (www.canrisk.org, CanRisk version 2).

Additional Information

Acknowledgements

Authors' contributions

Conceptualization: ACA, PPDP, DFE

Methodology: AL, ACA, DFE, PPDP

Software: AL, APC, TC, LF, SA, FMW, MT, JR, JUS

Investigation: AL, ACA, PPDP, DFE

Formal analysis: AL, NM, LF

Funding acquisition: ACA, MKS, PD, JS

Resources: PPDP, DFE, VZ, HJ, EF, AdP, MR, TM

Supervision: ACA

Visualisation: AL

Writing – original draft: AL, ACA, DFE

Writing – review & editing: all authors

Ethics approval and consent to participate

No ethical approval or consent was necessary for this work.

Data availability

Both models are freely available for non-commercial use through www.canrisk.org

Competing interests

The authors ACA, DFE, AL, APC and TC are named inventors of BOADICEA v5 commercialised by Cambridge Enterprise.

Funding information

This work has been supported by grants from Cancer Research UK (C12292/A20861 and PPRPGM-Nov20\100002); 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 the Government of Canada funds 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. MT is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014).

The University of Cambridge receives salary support for PDPP from the NHS through the Clinical Academic Reserve.

This research is linked to the CanTest Collaborative, which is funded by Cancer Research UK [C8640/A23385], of which FMW is Director.

References

1. Cancer Research UK. Ovarian Cancer Statistics [Ovarian cancer statistics for the UK]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer accessed 21 June 2021.
2. Cancer Research UK. Breast cancer statistics [Breast cancer statistics for the UK]. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer accessed 21 June 2021.
3. The UK Government. Breast screening: programme overview [updated 17 March 2021. Available from: https://www.gov.uk/guidance/breast-screening-programme-overview accessed 15 July 2021.
4. Pashayan N, Antoniou AC, Ivanus U, Esserman LJ, Easton DF, French D, Sroczynski G, Hall P, Cuzick J, Evans DG, Simard J, Garcia-Closas M, Schmutzler R, Wegwarth O, Pharoah P, Moorthie S, De Montgolfier S, Baron C, Herceg Z, Turnbull C, Balleyguier C, Rossi PG, Wesseling J, Ritchie D, Tischkowitz M, Broeders M, Reisel D, Metspalu A, Callender T, de Koning H, Devilee P, Delaloge S, Schmidt MK, Widschwendter M. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol* 2020;17(11):687-705. doi: 10.1038/s41571-020-0388-9 [published Online First: 2020/06/20]
5. Maas P, Barrdahl M, Joshi AD, Auer PL, Gaudet MM, Milne RL, Schumacher FR, Anderson WF, Check D, Chattopadhyay S, Baglietto L, Berg CD, Chanock SJ, Cox DG, Figueroa JD, Gail MH, Graubard BI, Haiman CA, Hankinson SE, Hoover RN, Isaacs C, Kolonel LN, Le Marchand L, Lee IM, Lindstrom S, Overvad K, Romieu I, Sanchez MJ, Southey MC, Stram DO, Tumino R, VanderWeele TJ, Willett WC, Zhang S, Buring JE, Canzian F, Gapstur SM, Henderson BE, Hunter DJ, Giles GG, Prentice RL, Ziegler RG, Kraft P, Garcia-Closas M, Chatterjee N. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol* 2016;2(10):1295-302. doi: 10.1001/jamaoncol.2016.1025
6. Pearce CL, Rossing MA, Lee AW, Ness RB, Webb PM, for Australian Cancer S, Australian Ovarian Cancer Study G, Chenevix-Trench G, Jordan SM, Stram DA, Chang-Claude J, Hein R, Nickels S, Lurie G, Thompson PJ, Carney ME, Goodman MT, Moysich K, Hogdall E, Jensen A, Goode EL, Fridley BL, Cunningham JM, Vierkant RA, Weber RP, Ziogas A, Anton-Culver H, Gayther SA, Gentry-Maharaj A, Menon U, Ramus SJ, Brinton L, Wentzensen N, Lissowska J, Garcia-Closas M, Massuger LF, Kiemeney LA, Van Altena AM, Aben KK, Berchuck A, Doherty JA, Iversen E, McGuire V, Moorman PG, Pharoah P, Pike MC, Risch H, Sieh W, Stram DO, Terry KL, Whittemore A, Wu AH, Schildkraut JM, Kjaer SK, Ovarian Cancer Association C. Combined and interactive effects of environmental and GWAS-identified risk factors in ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2013;22(5):880-90. doi: 10.1158/1055-9965.EPI-12-1030-T [published Online First: 2013/03/07]
7. Pearce CL, Stram DO, Ness RB, Stram DA, Roman LD, Templeman C, Lee AW, Menon U, Fasching PA, McAlpine JN, Doherty JA, Modugno F, Schildkraut JM, Rossing MA, Huntsman DG, Wu AH, Berchuck A, Pike MC, Pharoah PD. Population distribution of lifetime risk of ovarian cancer in the United States. *Cancer Epidemiol Biomarkers Prev* 2015;24(4):671-6. doi: 10.1158/1055-9965.EPI-14-1128
8. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, Babb de Villiers C, Izquierdo A, Simard J, Schmidt MK, Walter FM, Chatterjee N, Garcia-Closas M,

Tischkowitz M, Pharoah P, Easton DF, Antoniou AC. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med* 2019;21(8):1708-18. doi: 10.1038/s41436-018-0406-9 [published Online First: 2019/01/16]

9. Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA, Risch HA, Eyfjord JE, Hopper JL, Southey MC, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tryggvadottir L, Syrjakoski K, Kallioniemi OP, Eerola H, Nevanlinna H, Pharoah PD, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* 2008;98(8):1457-66. doi: 6604305 [pii];10.1038/sj.bjc.6604305 [doi]
10. Jervis S, Song H, Lee A, Dicks E, Harrington P, Baynes C, Manchanda R, Easton DF, Jacobs I, Pharoah PP, Antoniou AC. A risk prediction algorithm for ovarian cancer incorporating BRCA1, BRCA2, common alleles and other familial effects. *J Med Genet* 2015;52(7):465-75. doi: 10.1136/jmedgenet-2015-103077
11. Lee A, Yang X, Tyrer J, Gentry-Maharaj A, Ryan A, Mavaddat N, Cunningham AP, Carver T, Archer S, Leslie G, Kalsi J, Gaba F, Manchanda R, Gayther SA, Ramus SJ, Walter FM, Tischkowitz M, Jacobs I, Menon U, Easton DF, Pharoah PPD, Antoniou AC. A Comprehensive Epithelial Tubo-Ovarian Cancer Risk Prediction Model Incorporating Genetic and Epidemiological Risk Factors. *J Med Genet* 2021:In Press.
12. Lee AJ, Cunningham AP, Kuchenbaecker KB, Mavaddat N, Easton DF, Antoniou AC, Consortium of Investigators of Modifiers of B, Breast Cancer Association C. BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *Br J Cancer* 2014;110(2):535-45. doi: 10.1038/bjc.2013.730
13. Fischer C, Kuchenbacker K, Engel C, Zachariae S, Rhiem K, Meindl A, Rahner N, Dikow N, Plendl H, Debatin I, Grimm T, Gadzicki D, Flottmann R, Horvath J, Schrock E, Stock F, Schafer D, Schwaab I, Kartsonaki C, Mavaddat N, Schlegelberger B, Antoniou AC, Schmutzler R. Evaluating the performance of the breast cancer genetic risk models BOADICEA, IBIS, BRCAPRO and Claus for predicting BRCA1/2 mutation carrier probabilities: a study based on 7352 families from the German Hereditary Breast and Ovarian Cancer Consortium. *J Med Genet* 2013;50(6):360-67. doi: jmedgenet-2012-101415 [pii];10.1136/jmedgenet-2012-101415 [doi]
14. Carver T, Hartley S, Lee A, Cunningham AP, Archer S, Babb de Villiers C, Roberts J, Ruston R, Walter FM, Tischkowitz M, Easton DF, Antoniou AC. 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(3):469-73. doi: 10.1158/1055-9965.EPI-20-1319 [published Online First: 2020/12/19]
15. Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, Pooley KA, Parsons MT, Fortuno C, Wang Q, Bolla MK, Dennis J, Keeman R, Alonso MR, Álvarez N, Herraez B, Fernandez V, Núñez-Torres R, Osorio A, Valcich J, Li M, Törngren T, Harrington PA, Baynes C, Conroy DM, Decker B, Fachal L, Mavaddat N, Ahearn T, Aittomäki K, Antonenkova NN, Arnold N, Arveux P, Ausems MGEM, Auvinen P, Becher H, Beckmann MW, Behrens S, Bermisheva M, Białkowska K, Blomqvist C, Bogdanova NV, Bogdanova-Markov N, Bojesen SE, Bonanni B, Børresen-Dale AL, Brauch H, Bremer M, Briceno I, Brüning T, Burwinkel B, Cameron DA, Camp NJ, Campbell A, Carracedo A, Castelao JE, Cessna MH, Chanock SJ, Christiansen H, Collée

JM, Cordina-Duverger E, Cornelissen S, Czene K, Dörk T, Ekici AB, Engel C, Eriksson M, Fasching PA, Figueroa J, Flyger H, Försti A, Gabrielson M, Gago-Dominguez M, Georgoulias V, Gil F, Giles GG, Glendon G, Garcia EBG, Alnæs GIG, Guénel P, Hadjisavvas A, Haeberle L, Hahnen E, Hall P, Hamann U, Harkness EF, Hartikainen JM, Hartman M, He W, Heemskerk-Gerritsen BAM, Hillemanns P, Hogervorst FBL, Hollestelle A, Ho WK, Hooning MJ, Howell A, Humphreys K, Idris F, Jakubowska A, Jung A, Kapoor PM, Kerin MJ, Khusnutdinova E, Kim SW, Ko YD, Kosma VM, Kristensen VN, Kyriacou K, Lakeman IMM, Lee JW, Lee MH, Li J, Lindblom A, Lo WY, Loizidou MA, Lophatananon A, Lubiński J, MacInnis RJ, Madsen MJ, Mannermaa A, Manoochehri M, Manoukian S, Margolin S, Martinez ME, Maurer T, Mavroudis D, McLean C, Meindl A, Mensenkamp AR, Michailidou K, Miller N, Mohd Taib NA, Muir K, Mulligan AM, Nevanlinna H, Newman WG, Nordestgaard BG, Ng PS, Oosterwijk JC, Park SK, Park-Simon TW, Perez JIA, Peterlongo P, Porteous DJ, Prajzendanc K, Prokofyeva D, Radice P, Rashid MU, Rhenius V, Rookus MA, Rüdiger T, Saloustros E, Sawyer EJ, Schmutzler RK, Schneeweiss A, Schürmann P, Shah M, Sohn C, Southey MC, Surowy H, Suvanto M, Thanasitthichai S, Tomlinson I, Torres D, Truong T, Tzardi M, Valova Y, van Asperen CJ, Van Dam RM, van den Ouweland AMW, van der Kolk LE, van Veen EM, Wendt C, Williams JA, Yang XR, Yoon SY, Zamora MP, Evans DG, de la Hoya M, Simard J, Antoniou AC, Borg Å, Andrulis IL, Chang-Claude J, García-Closas M, Chenevix-Trench G, Milne RL, Pharoah PDP, Schmidt MK, Spurdle AB, Vreeswijk MPG, Benitez J, Dunning AM, Kvist A, Teo SH, Devilee P, Easton DF, Consortium BCA. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* 2021;384(5):428-39. doi: 10.1056/NEJMoa1913948 [published Online First: 2021/01/20]

16. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, Gao C, Lilyquist J, Yadav S, Boddicker NJ, Samara R, Klebba J, Ambrosone CB, Anton-Culver H, Auer P, Bandera EV, Bernstein L, Bertrand KA, Burnside ES, Carter BD, Eliassen H, Gapstur SM, Gaudet M, Haiman C, Hodge JM, Hunter DJ, Jacobs EJ, John EM, Kooperberg C, Kurian AW, Le Marchand L, Lindstroem S, Lindstrom T, Ma H, Neuhausen S, Newcomb PA, O'Brien KM, Olson JE, Ong IM, Pal T, Palmer JR, Patel AV, Reid S, Rosenberg L, Sandler DP, Scott C, Tamimi R, Taylor JA, Trentham-Dietz A, Vachon CM, Weinberg C, Yao S, Ziogas A, Weitzel JN, Goldgar DE, Domchek SM, Nathanson KL, Kraft P, Polley EC, Couch FJ. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *N Engl J Med* 2021;384(5):440-51. doi: 10.1056/NEJMoa2005936 [published Online First: 2021/01/21]

17. Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, Dunning AM, Redman J, Scarth J, Plaskocinska I, Luccarini C, Shah M, Pooley K, Dorling L, Lee A, Adank MA, Adlard J, Aittomaki K, Andrulis IL, Ang P, Barwell J, Bernstein JL, Bobolis K, Borg A, Blomqvist C, Claes KBM, Concannon P, Cuggia A, Culver JO, Damiola F, de Pauw A, Diez O, Dolinsky JS, Domchek SM, Engel C, Evans DG, Fostira F, Garber J, Golmard L, Goode EL, Gruber SB, Hahnen E, Hake C, Heikkinen T, Hurley JE, Janavicius R, Kleibl Z, Kleiblova P, Konstantopoulou I, Kvist A, Laduca H, Lee ASG, Lesueur F, Maher ER, Mannermaa A, Manoukian S, McFarland R, McKinnon W, Meindl A, Metcalfe K, Mohd Taib NA, Moilanen J, Nathanson KL, Neuhausen S, Ng PS, Nguyen-Dumont T, Nielsen SM, Obermair F, Offit K, Olopade OI, Ottini L, Penkert J, Pylkas K, Radice P, Ramus SJ, Rudaitis V, Side L, Silva-Smith R, Silvestri V, Skytte AB, Slavin T, Soukupova J, Tondini C, Trainer AH, Unzeitig G, Usha L, van Overeem Hansen T, Whitworth J,

Wood M, Yip CH, Yoon SY, Yussuf A, Zogopoulos G, Goldgar D, Hopper JL, Chenevix-Trench G, Pharoah P, George SHL, Balmana J, Houdayer C, James P, El-Haffaf Z, Ehrencrona H, Janatova M, Peterlongo P, Nevanlinna H, Schmutzler R, Teo SH, Robson M, Pal T, Couch F, Weitzel JN, Elliott A, Southey M, Winqvist R, Easton DF, Foulkes WD, Antoniou AC, Tischkowitz M. Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *J Clin Oncol* 2020;38(7):674-85. doi: 10.1200/JCO.19.01907 [published Online First: 2019/12/17]

18. Song H, Dicks EM, Tyrer J, Intermaggio M, Chenevix-Trench G, Bowtell DD, Traficante N, Group A, Brenton J, Goranova T, Hosking K, Piskorz A, van Oudenhove E, Doherty J, Harris HR, Rossing MA, Duerst M, Dork T, Bogdanova NV, Modugno F, Moysich K, Odunsi K, Ness R, Karlan BY, Lester J, Jensen A, Kruger Kjaer S, Hogdall E, Campbell IG, Lazaro C, Pujara MA, Cunningham J, Vierkant R, Winham SJ, Hildebrandt M, Huff C, Li D, Wu X, Yu Y, Permuth JB, Levine DA, Schildkraut JM, Riggan MJ, Berchuck A, Webb PM, Group OS, Cybulski C, Gronwald J, Jakubowska A, Lubinski J, Alsop J, Harrington P, Chan I, Menon U, Pearce CL, Wu AH, de Fazio A, Kennedy CJ, Goode E, Ramus S, Gayther S, Pharoah P. Population-based targeted sequencing of 54 candidate genes identifies PALB2 as a susceptibility gene for high-grade serous ovarian cancer. *J Med Genet* 2021;58(5):305-13. doi: 10.1136/jmedgenet-2019-106739 [published Online First: 2020/06/18]
19. Breast Cancer Association C, Mavaddat N, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Keeman R, Bolla MK, Dennis J, Wang Q, Ahearn TU, Andrulis IL, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Blomqvist C, Bogdanova NV, Bojesen SE, Briceno I, Bruning T, Camp NJ, Campbell A, Castelao JE, Chang-Claude J, Chanock SJ, Chenevix-Trench G, Christiansen H, Czene K, Dork T, Eriksson M, Evans DG, Fasching PA, Figueroa JD, Flyger H, Gabrielson M, Gago-Dominguez M, Geisler J, Giles GG, Guenel P, Hadjisavvas A, Hahnen E, Hall P, Hamann U, Hartikainen JM, Hartman M, Hoppe R, Howell A, Jakubowska A, Jung A, Khusnutdinova EK, Kristensen VN, Li J, Lim SH, Lindblom A, Loizidou MA, Lophatananon A, Lubinski J, Madsen MJ, Mannermaa A, Manoochehri M, Margolin S, Mavroudis D, Milne RL, Mohd Taib NA, Morra A, Muir K, Obi N, Osorio A, Park-Simon TW, Peterlongo P, Radice P, Saloustros E, Sawyer EJ, Schmutzler RK, Shah M, Sim X, Southey MC, Thorne H, Tomlinson I, Torres D, Truong T, Yip CH, Spurdle AB, Vreeswijk MPG, Dunning AM, Garcia-Closas M, Pharoah PDP, Kvist A, Muranen TA, Nevanlinna H, Teo SH, Devilee P, Schmidt MK, Easton DF. Pathology of Tumors Associated With Pathogenic Germline Variants in 9 Breast Cancer Susceptibility Genes. *JAMA Oncol* 2022;8(3):e216744. doi: 10.1001/jamaoncol.2021.6744 [published Online First: 2022/01/28]
20. Mavaddat N, Rebbeck TR, Lakhani SR, Easton DF, Antoniou AC. Incorporating tumour pathology information into breast cancer risk prediction algorithms. *Breast Cancer Res* 2010;12(3):R28. doi: bcr2576 [pii];10.1186/bcr2576 [doi]
21. Lee AJ, Cunningham AP, Tischkowitz M, Simard J, Pharoah PD, Easton DF, Antoniou AC. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. *Genet Med* 2016 doi: 10.1038/gim.2016.31
22. Antoniou AC, Pharoah PP, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer* 2004;91(8):1580-90. doi: 10.1038/sj.bjc.6602175 [doi];6602175 [pii]

23. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst* 2014;106(11):dju305. doi: 10.1093/jnci/dju305
24. Yang X, Song H, Leslie G, Engel C, Hahnen E, Auber B, Horvath J, Kast K, Niederacher D, Turnbull C, Houlston R, Hanson H, Loveday C, Dolinsky JS, LaDuca H, Ramus SJ, Menon U, Rosenthal AN, Jacobs I, Gayther SA, Dicks E, Nevanlinna H, Aittomaki K, Pelttari LM, Ehrencrona H, Borg A, Kvist A, Rivera B, Hansen TVO, Djursby M, Lee A, Dennis J, Bowtell DD, Traficante N, Diez O, Balmana J, Gruber SB, Chenevix-Trench G, kConFab I, Jensen A, Kjaer SK, Hogdall E, Castera L, Garber J, Janavicius R, Osorio A, Golmard L, Vega A, Couch FJ, Robson M, Gronwald J, Domchek SM, Culver JO, de la Hoya M, Easton DF, Foulkes WD, Tischkowitz M, Meindl A, Schmutzler RK, Pharoah PDP, Antoniou AC. Ovarian and breast cancer risks associated with pathogenic variants in RAD51C and RAD51D. *J Natl Cancer Inst* 2020 doi: 10.1093/jnci/djaa030 [published Online First: 2020/02/29]
25. National Institute for Health and Care Excellence. Familial breast cancer: classification and care of people at risk of familial breast cancer and management of breast cancer and related risks in people with a family history of breast cancer (CG164) 2013 [updated 2019-11-20. Available from: http://www.nice.org.uk/CG164.
26. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen TH, Wang Q, Bolla MK, Yang X, Adank MA, Ahearn T, Aittomaki K, Allen J, Andrulis IL, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Auer PL, Auvinen P, Barrdahl M, Beane Freeman LE, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Borresen-Dale AL, Brauch H, Bremer M, Brenner H, Brentnall A, Brock IW, Brooks-Wilson A, Brucker SY, Bruning T, Burwinkel B, Campa D, Carter BD, Castelao JE, Chanock SJ, Chlebowski R, Christiansen H, Clarke CL, Collee JM, Cordina-Duverger E, Cornelissen S, Couch FJ, Cox A, Cross SS, Czene K, Daly MB, Devilee P, Dork T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Engel C, Eriksson M, Evans DG, Fasching PA, Figueroa J, Fletcher O, Flyger H, Forsti A, Fritschi L, Gabrielson M, Gago-Dominguez M, Gapstur SM, Garcia-Saenz JA, Gaudet MM, Georgoulias V, Giles GG, Gilyazova IR, Glendon G, Goldberg MS, Goldgar DE, Gonzalez-Neira A, Grenaker Alnaes GI, Grip M, Gronwald J, Grundy A, Guenel P, Haeberle L, Hahnen E, Haiman CA, Hakansson N, Hamann U, Hankinson SE, Harkness EF, Hart SN, He W, Hein A, Heyworth J, Hillemanns P, Hollestelle A, Hooning MJ, Hoover RN, Hopper JL, Howell A, Huang G, Humphreys K, Hunter DJ, Jakimovska M, Jakubowska A, Janni W, John EM, Johnson N, Jones ME, Jukkola-Vuorinen A, Jung A, Kaaks R, Kaczmarek K, Kataja V, Keeman R, Kerin MJ, Khusnutdinova E, Kiiski JI, Knight JA, Ko YD, Kosma VM, Koutros S, Kristensen VN, Kruger U, Kuhl T, Lambrechts D, Le Marchand L, Lee E, Lejbkowicz F, Lilyquist J, Lindblom A, Lindstrom S, Lissowska J, Lo WY, Loibl S, Long J, Lubinski J, Lux MP, MacInnis RJ, Maishman T, Makalic E, Maleva Kostovska I, Mannermaa A, Manoukian S, Margolin S, Martens JWM, Martinez ME, Mavroudis D, McLean C, Meindl A, Menon U, Middha P, Miller N, Moreno F, Mulligan AM, Mulot C, Munoz-Garzon VM, Neuhausen SL, Nevanlinna H, Neven P, Newman WG, Nielsen SF, Nordestgaard BG, Norman A, Offit K, Olson JE, Olsson H, Orr N, Pankratz VS, Park-Simon TW, Perez JIA, Perez-Barrios C, Peterlongo P, Peto J, Pinchev M, Plaseska-Karanfilska D, Polley EC, Prentice R, Presneau N, Prokofyeva D, Purrington K, Pylkas K, Rack B, Radice P, Rau-Murthy R, Rennert G,

Rennert HS, Rhenius V, Robson M, Romero A, Ruddy KJ, Ruebner M, Saloustros E, Sandler DP, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schurmann P, Schwentner L, Scott C, Scott RJ, Seynaeve C, Shah M, Sherman ME, Shrubsole MJ, Shu XO, Slager S, Smeets A, Sohn C, Soucy P, Southey MC, Spinelli JJ, Stegmaier C, Stone J, Swerdlow AJ, Tamimi RM, Tapper WJ, Taylor JA, Terry MB, Thone K, Tollenaar R, Tomlinson I, Truong T, Tzardi M, Ulmer HU, Untch M, Vachon CM, van Veen EM, Vijai J, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Yang XR, Yannoukakos D, Zhang Y, Zheng W, Ziogas A, Investigators A, kConFab AI, Collaborators N, Dunning AM, Thompson DJ, Chenevix-Trench G, Chang-Claude J, Schmidt MK, Hall P, Milne RL, Pharoah PDP, Antoniou AC, Chatterjee N, Kraft P, Garcia-Closas M, Simard J, Easton DF. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2019;104(1):21-34. doi: 10.1016/j.ajhg.2018.11.002

27. Manchanda R, Legood R, Antoniou AC, Gordeev VS, Menon U. Specifying the ovarian cancer risk threshold of 'premenopausal risk-reducing salpingo-oophorectomy' for ovarian cancer prevention: a cost-effectiveness analysis. *J Med Genet* 2016;53(9):591-9. doi: 10.1136/jmedgenet-2016-103800
28. Manchanda R, Legood R, Pearce L, Menon U. Defining the risk threshold for risk reducing salpingo-oophorectomy for ovarian cancer prevention in low risk postmenopausal women. *Gynecol Oncol* 2015;139(3):487-94. doi: 10.1016/j.ygyno.2015.10.001
29. Dareng EO, Tyrer J, Barnes DR, Jones MR, Yang X, Aben KKH, Adank MA, Agata SA, Andrulis IL, Anton-Culver H, Antonenkova NN, Aravantinos G, Arun BK, Augustinsson A, Balmaña J, Bandera EV, Barkardottir RB, Barrowdale D, Beckmann MW, Beeghly-Fadiel A, Benitez J, Bermisheva M, Bernardini MQ, Bjorge L, Black A, Bogdanova NV, Bonanni B, Borg A, Brenton J, Budzilowska A, Butzow R, Buys SS, Cai H, Caligo MA, Campbell I, Cannioto R, Cassingham H, Chang-Claude J, Chanock SJ, Chen K, Chiew Y-E, Chung WK, Claes KBM, Colanna S, Cook LS, Couch FJ, Daly MB, Dao F, Davies E, de la Hoya M, de Putter R, DePersia A, Devilee P, Diez O, Ding YC, Doherty JA, Domchek SM, Dörk T, du Bois A, Dürst M, Eccles DM, Eliassen HA, Engel C, Evans DG, Fasching P, Flanagan JM, Foretova L, Fortner RT, Friedman E, Ganz PA, Garber J, Gensini F, Giles GG, Glendon G, Godwin AK, Goodman MT, Greene MH, Gronwald J, Hahnen E, Haiman CA, Håkansson N, Hamann U, Hansen TVO, Harris HR, Hartman M, Heitz F, Hildebrandt MAT, Høgdall E, Høgdall CK, Hopper JL, Huang R-Y, Huff C, Hulick PJ, Huntsman DG, Imyanitov EN, Isaacs C, Anna Jakubowska A, James P, Janavicius R, Jensen A, Johannsson OT, John EM, Jones M, Kang D, Karlan BY, Karnezis A, Kelemen LE, Khusnutdinova E, Kiemeney LA, Kim B-G, Kjaer SK, Komenaka I, Kupryjanczyk J, Kurian AW, Kwong A, Lambrechts D, Larson MC, Lazaro C, Le ND, Leslie G, Lester J, Lesueur F, Levine D, Li L, Li J, Loud JT, Lu KH, Lubiński J, Machackova E, Mai PL, Manoukian S, Marks J, Kim Matsuno R, Matsuo K, May T, McGuffog L, McLaughlin JR, McNeish IA, Mebirouk N, Menon U, Miller A, Milne RL, Minlikeeva A, Modugno F, Montagna M, Moysich KB, Munro E, Nathanson KL, Neuhausen SL, Nevanlinna H, Ngeow Yuen Yie J, Roed Nielsen H, Nielsen FC, Nikitina-Zake L, Odunsi K, Offit K, Olah E, Olbrecht S, Olopade OI, Olson SH, Olsson H, Osorio A, Papi L, Park SK, Parsons MT, Pathak H, Pedersen IS, Peixoto A, Pejovic T, Perez-Segura P, Permuth JB, Peshkin B, Peterlongo P, Piskorz A, Prokofyeva D, Radice P, Rantala J, Riggan M, Risch HB, Rodriguez-Antona C, Ross EA, Rossing MA, Runnebaum I, Sandler DP, Santamariña M, Soucy P, Schmutzler RK, Setiawan VW, Shan K, Sieh W, Simard J, Singer CF,

Sokolenko A, Song H, Southey M, Steed H, Stoppa-Lyonnet D, Sutphen R, Swerdlow AJ, Tan YY, Teixeira M, Teo SH, Terry KL, Terry MB, Thomassen M, Thompson PJ, Thomsen LC, Hull DL, Tischkowitz M, Titus L, Toland A, Torres D, Trabert B, Travis R, Tung N, Tworoger SS, Valen E, van Altena AM, van der Hout AH, Van Nieuwenhuysen E, van Rensburg EJ, Vega A, Velez Edwards D, Vierkant RA, Wang F, Wappenschmidt B, Webb PM, Weinberg CR, Weitzel JN, Wentzensen N, White E, Whittemore AS, Winham SJ, Wolk A, Woo Y-L, Wu AH, Yan L, Yannoukakos D, Zavaglia KM, Zheng W, Ziogas A, Zorn KK, Easton D, Lawrenson K, deFazio A, Sellers TA, Ramus SJ, Pearce C, Montiero AN, Cunningham JM, Goode EL, Schildkraut JM, Berchuck A, Chenevix-Trench G, Gayther SA, Antoniou AC, Pharoah PDP. Polygenic Risk Modelling for Prediction of Epithelial Ovarian Cancer Risk. *Eur J Hum Genet* 2021:In Press.

30. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol* 2001;21(1):1-18. doi: 10.1002/gepi.1014 [pii];10.1002/gepi.1014 [doi]
31. Muranen TA, Greco D, Blomqvist C, Aittomaki K, Khan S, Hogervorst F, Verhoef S, Pharoah PDP, Dunning AM, Shah M, Luben R, Bojesen SE, Nordestgaard BG, Schoemaker M, Swerdlow A, Garcia-Closas M, Figueroa J, Dork T, Bogdanova NV, Hall P, Li J, Khusnutdinova E, Bermisheva M, Kristensen V, Borresen-Dale AL, Investigators N, Peto J, Dos Santos Silva I, Couch FJ, Olson JE, Hillemans P, Park-Simon TW, Brauch H, Hamann U, Burwinkel B, Marme F, Meindl A, Schmutzler RK, Cox A, Cross SS, Sawyer EJ, Tomlinson I, Lambrechts D, Moisse M, Lindblom A, Margolin S, Hollestelle A, Martens JWM, Fasching PA, Beckmann MW, Andrulis IL, Knight JA, Investigators K, Anton-Culver H, Ziogas A, Giles GG, Milne RL, Brenner H, Arndt V, Mannermaa A, Kosma VM, Chang-Claude J, Rudolph A, Devilee P, Seynaeve C, Hopper JL, Southey MC, John EM, Whittemore AS, Bolla MK, Wang Q, Michailidou K, Dennis J, Easton DF, Schmidt MK, Nevanlinna H. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. *Genet Med* 2017;19(5):599-603. doi: 10.1038/gim.2016.147
32. Gao C, Polley EC, Hart SN, Huang H, Hu C, Gnanaolivu R, Lilyquist J, Boddicker NJ, Na J, Ambrosone CB, Auer PL, Bernstein L, Burnside ES, Eliassen AH, Gaudet MM, Haiman C, Hunter DJ, Jacobs EJ, John EM, Lindstrom S, Ma H, Neuhausen SL, Newcomb PA, O'Brien KM, Olson JE, Ong IM, Patel AV, Palmer JR, Sandler DP, Tamimi R, Taylor JA, Teras LR, Trentham-Dietz A, Vachon CM, Weinberg CR, Yao S, Weitzel JN, Goldgar DE, Domchek SM, Nathanson KL, Couch FJ, Kraft P. Risk of Breast Cancer Among Carriers of Pathogenic Variants in Breast Cancer Predisposition Genes Varies by Polygenic Risk Score. *J Clin Oncol* 2021:JCO2001992. doi: 10.1200/JCO.20.01992 [published Online First: 2021/06/09]
33. Gallagher S, Hughes E, Wagner S, Tshiaba P, Rosenthal E, Roa BB, Kurian AW, Domchek SM, Garber J, Lancaster J, Weitzel JN, Gutin A, Lanchbury JS, Robson M. Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. *JAMA Netw Open* 2020;3(7):e208501. doi: 10.1001/jamanetworkopen.2020.8501 [published Online First: 2020/07/02]
34. Kapoor PM, Mavaddat N, Choudhury PP, Wilcox AN, Lindstrom S, Behrens S, Michailidou K, Dennis J, Bolla MK, Wang Q, Jung A, Abu-Ful Z, Ahearn T, Andrulis IL, Anton-Culver H, Arndt V, Aronson KJ, Auer PL, Freeman LEB, Becher H, Beckmann MW, Beeghly-Fadiel A, Benitez J, Bernstein L, Bojesen SE, Brauch H, Brenner H, Bruning T, Cai Q, Campa D, Canzian F, Carracedo A, Carter BD, Castelao JE, Chanock SJ, Chatterjee N,

Chenevix-Trench G, Clarke CL, Couch FJ, Cox A, Cross SS, Czene K, Dai JY, Earp HS, Ekici AB, Eliassen AH, Eriksson M, Evans DG, Fasching PA, Figueroa J, Fritschi L, Gabrielson M, Gago-Dominguez M, Gao C, Gapstur SM, Gaudet MM, Giles GG, Gonzalez-Neira A, Guenel P, Haeberle L, Haiman CA, Hakansson N, Hall P, Hamann U, Hatse S, Heyworth J, Holleczek B, Hoover RN, Hopper JL, Howell A, Hunter DJ, Investigators A, kConFab AI, John EM, Jones ME, Kaaks R, Keeman R, Kitahara CM, Ko YD, Koutros S, Kurian AW, Lambrechts D, Le Marchand L, Lee E, Lejbkowicz F, Linet M, Lissowska J, Llaneza A, MacInnis RJ, Martinez ME, Maurer T, McLean C, Neuhausen SL, Newman WG, Norman A, O'Brien KM, Olshan AF, Olson JE, Olsson H, Orr N, Perou CM, Pita G, Polley EC, Prentice RL, Rennert G, Rennert HS, Ruddy KJ, Sandler DP, Saunders C, Schoemaker MJ, Schottker B, Schumacher F, Scott C, Scott RJ, Shu XO, Smeets A, Southey MC, Spinelli JJ, Stone J, Swerdlow AJ, Tamimi RM, Taylor JA, Troester MA, Vachon CM, van Veen EM, Wang X, Weinberg CR, Weltens C, Willett W, Winham SJ, Wolk A, Yang XR, Zheng W, Ziogas A, Dunning AM, Pharoah PDP, Schmidt MK, Kraft P, Easton DF, Milne RL, Garcia-Closas M, Chang-Claude J. Combined Associations of a Polygenic Risk Score and Classical Risk Factors With Breast Cancer Risk. *J Natl Cancer Inst* 2021;113(3):329-37. doi: 10.1093/jnci/djaa056 [published Online First: 2020/05/03]

35. Schmidt MK, Hogervorst F, van Hien R, Cornelissen S, Broeks A, Adank MA, Meijers H, Waisfisz Q, Hollestelle A, Schutte M, van den Ouweland A, Hooning M, Andrulis IL, Anton-Culver H, Antonenkova NN, Antoniou AC, Arndt V, Bermisheva M, Bogdanova NV, Bolla MK, Brauch H, Brenner H, Bruning T, Burwinkel B, Chang-Claude J, Chenevix-Trench G, Couch FJ, Cox A, Cross SS, Czene K, Dunning AM, Fasching PA, Figueroa J, Fletcher O, Flyger H, Galle E, Garcia-Closas M, Giles GG, Haeberle L, Hall P, Hillemanns P, Hopper JL, Jakubowska A, John EM, Jones M, Khusnutdinova E, Knight JA, Kosma VM, Kristensen V, Lee A, Lindblom A, Lubinski J, Mannermaa A, Margolin S, Meindl A, Milne RL, Muranen TA, Newcomb PA, Offit K, Park-Simon TW, Peto J, Pharoah PD, Robson M, Rudolph A, Sawyer EJ, Schmutzler RK, Seynaeve C, Soens J, Southey MC, Spurdle AB, Surowy H, Swerdlow A, Tollenaar RA, Tomlinson I, Trentham-Dietz A, Vachon C, Wang Q, Whittemore AS, Ziogas A, van der Kolk L, Nevanlinna H, Dork T, Bojesen S, Easton DF. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *J Clin Oncol* 2016;34(23):2750-60. doi: 10.1200/JCO.2016.66.5844
36. Eriksson M, Li J, Leifland K, Czene K, Hall P. A comprehensive tool for measuring mammographic density changes over time. *Breast cancer research and treatment* 2018;169(2):371-79. doi: 10.1007/s10549-018-4690-5 [published Online First: 2018/02/03]
37. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. *Phys Med Biol* 1994;39(10):1629-38. doi: 10.1088/0031-9155/39/10/008 [published Online First: 1994/10/01]
38. Astley SM, Harkness EF, Sergeant JC, Warwick J, Stavrinos P, Warren R, Wilson M, Beetles U, Gadde S, Lim Y, Jain A, Bundred S, Barr N, Reece V, Brentnall AR, Cuzick J, Howell T, Evans DG. A comparison of five methods of measuring mammographic density: a case-control study. *Breast Cancer Res* 2018;20(1):10. doi: 10.1186/s13058-018-0932-z [published Online First: 2018/02/07]

39. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* 2016;17(7):392-406. doi: 10.1038/nrg.2016.27
40. Pal Choudhury P, Maas P, Wilcox A, Wheeler W, Brook M, Check D, Garcia-Closas M, Chatterjee N. iCARE: An R package to build, validate and apply absolute risk models. *PLoS One* 2020;15(2):e0228198. doi: 10.1371/journal.pone.0228198 [published Online First: 2020/02/05]
41. Pal Choudhury P, Brook MN, Hurson AN, Lee A, Mulder CV, Coulson P, Schoemaker MJ, Jones ME, Swerdlow AJ, Chatterjee N, Antoniou AC, Garcia-Closas M. 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(1):22. doi: 10.1186/s13058-021-01399-7 [published Online First: 2021/02/15]
42. Pal Choudhury P, Wilcox AN, Brook MN, Zhang Y, Ahearn T, Orr N, Coulson P, Schoemaker MJ, Jones ME, Gail MH, Swerdlow AJ, Chatterjee N, Garcia-Closas M. Comparative Validation of Breast Cancer Risk Prediction Models and Projections for Future Risk Stratification. *J Natl Cancer Inst* 2020;112(3):278-85. doi: 10.1093/jnci/djz113 [published Online First: 2019/06/06]
43. Ho WK, Tan MM, Mavaddat N, Tai MC, Mariapun S, Li J, Ho PJ, Dennis J, Tyrer JP, Bolla MK, Michailidou K, Wang Q, Kang D, Choi JY, Jamaris S, Shu XO, Yoon SY, Park SK, Kim SW, Shen CY, Yu JC, Tan EY, Chan PMY, Muir K, Lophatananon A, Wu AH, Stram DO, Matsuo K, Ito H, Chan CW, Ngeow J, Yong WS, Lim SH, Lim GH, Kwong A, Chan TL, Tan SM, Seah J, John EM, Kurian AW, Koh WP, Khor CC, Iwasaki M, Yamaji T, Tan KMV, Tan KTB, Spinelli JJ, Aronson KJ, Hasan SN, Rahmat K, Vijayananthan A, Sim X, Pharoah PDP, Zheng W, Dunning AM, Simard J, van Dam RM, Yip CH, Taib NAM, Hartman M, Easton DF, Teo SH, Antoniou AC. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun* 2020;11(1):3833. doi: 10.1038/s41467-020-17680-w [published Online First: 2020/08/02]
44. Shieh Y, Fejerman L, Lott PC, Marker K, Sawyer SD, Hu D, Huntsman S, Torres J, Echeverry M, Bohorquez ME, Martinez-Chequer JC, Polanco-Echeverry G, Estrada-Florez AP, Consortium C, Haiman CA, John EM, Kushi LH, Torres-Mejia G, Vidaurre T, Weitzel JN, Zambrano SC, Carvajal-Carmona LG, Ziv E, Neuhausen SL. A Polygenic Risk Score for Breast Cancer in US Latinas and Latin American Women. *J Natl Cancer Inst* 2020;112(6):590-98. doi: 10.1093/jnci/djz174 [published Online First: 2019/09/26]
45. Dorling L, Carvalho S, Allen J, Parsons MT, Fortuno C, González-Neira A, Heijl SM, Adank MA, Ahearn TU, Andrulis IL, Auvinen P, Becher H, Beckmann MW, Behrens S, Bermisheva M, Bogdanova NV, Bojesen SE, Bolla MK, Bremer M, Briceno I, Camp NJ, Campbell A, Castelao JE, Chang-Claude J, Chanock SJ, Chenevix-Trench G, Collaborators N, Collée JM, Czene K, Dennis J, Dörk T, Eriksson M, Evans DG, Fasching PA, Figueroa J, Flyger H, Gabrielson M, Gago-Dominguez M, García-Closas M, Giles GG, Glendon G, Guénel P, Gündert M, Hadjisavvas A, Hahnen E, Hall P, Hamann U, Harkness EF, Hartman M, Hogervorst FBL, Hollestelle A, Hoppe R, Howell A, Investigators k, Investigators S, Jakubowska A, Jung A, Khusnutdinova E, Kim S-W, Ko Y-D, Kristensen VN, Lakeman IMM, Li J, Lindblom A, Loizidou MA, Lophatananon A, Lubiński J, Luccarini C, Madsen MJ, Mannermaa A, Manoochehri M, Margolin S,

Mavroudis D, Milne RL, Mohd Taib NA, Muir K, Nevanlinna H, Newman WG, Oosterwijk JC, Park SK, Peterlongo P, Radice P, Saloustros E, Sawyer EJ, Schmutzler RK, Shah M, Sim X, Southey MC, Surowy H, Suvanto M, Tomlinson I, Torres D, Truong T, van Asperen CJ, Waltes R, Wang Q, Yang XR, Pharoah PDP, Schmidt MK, Benitez J, Vroling B, Dunning AM, Teo SH, Kvist A, de la Hoya M, Devilee P, Spurdle AB, Vreeswijk MPG, Easton DF. Breast cancer risks associated with missense variants in breast cancer susceptibility genes. *medRxiv* 2021:2021.09.02.21262369. doi: 10.1101/2021.09.02.21262369

Figure Legends

Figure 1: Predicted risk by age Vs age for a female born in 1985 with an unknown family history based on pathogenic variant carrier status for the new genes in the model. Figure (a) shows the breast cancer risk for carriers of pathogenic variants in *BARD1*, *RAD51C* and *RAD51D* along with the population risk. Figure (b) shows the ovarian cancer risk for carriers of pathogenic variants in *PALB2* along with the population risk. Predictions are based on UK cancer incidences.

Figure 2. Predicted lifetime cancer risks (from age 20 to 80 years) for a female born in 1985 with a pathogenic variant in *BARD1* (breast cancer risk), *RAD51C* (breast cancer risk), *RAD51D* (breast cancer risk), and *PALB2* (ovarian cancer risk) on the basis of the different predictors of risk (pathogenic variant status (PV), questionnaire-based risk factors (QRFs), mammographic density (MD), and PRS). All figures show the probability density against the absolute risk. Figures (a), (c), (e) and (g) show risks for a female with unknown family history, while Figures (b), (d), (f) and (h) show risks where the individual's mother has had cancer at age 50. The backgrounds of the graphs are shaded to indicate the risk categories. For breast cancer, these are the categories defined by the National Institute for Health and Care Excellence familial breast cancer guidelines[25]: 1) near-population risk shaded in pink (<17%), 2) moderate risk shaded in yellow (≥17% and <30%) and 3) high risk shaded in blue (≥30%). For ovarian cancer, the categories are: 1) near-population risk shaded in pink (<5%), 2) moderate risk shaded in yellow (≥5% and <10%) and 3) high risk shaded in blue (≥10%). Predictions were based on UK cancer incidences. The line labelled population denotes the average population risk in the absence of knowledge of family history, pathogenic variant status, RFs or a PRS. All figures assume the population distributions of QRFs and MD.

Figure 3. The tumour pathology proportions in the general population and among carriers of pathogenic variants (PV) in the breast cancer (BC) susceptibility genes included in the BOADICEA model. Figure (a) shows the proportion of oestrogen receptor-negative (ER-) tumours among all tumours, and Figure (b) shows the proportion of triple-negative (TN) (ER-, progesterone receptor-negative and human epidermal growth factor receptor 2) tumours among ER- tumours. The general population, *BRCA1* PV and *BRCA2* PV values are the same as

previously used in the model[12], while those for the other genes are updated using recent BRIDGES data[19].

Figure 4: The probabilities of carrying a pathogenic variant estimated by BOADICEA in the genes *PALB2*, *CHEK2*, *ATM*, *BARD1*, *RAD51C* and *RAD51D* for an affected female born in 1985 as a function of her age at diagnosis based on different tumour pathology. Figures (a), (c), (e) and (g) show the probabilities based on the updated proportions (current model), while figures (b), (d), (f) and (h) are based on the previously assumed tumour pathology proportions (previous model version) and where proportions for *BARD1*, *RAD51C* and *RAD51D*, which were not in the previous model, are assumed to be the same as in the general population. In figures (a) and (b), the female has had an oestrogen receptor-positive (ER+) tumour; in figures (c) and (d), the female has had an oestrogen receptor-negative (ER-) tumour, but the triple-negative (TN) status is unknown; in figures (e) and (f), the female has had an ER- tumour that is not TN and in figures (g) and (h), the female has had a TN tumour. Predictions are based on UK cancer incidences.

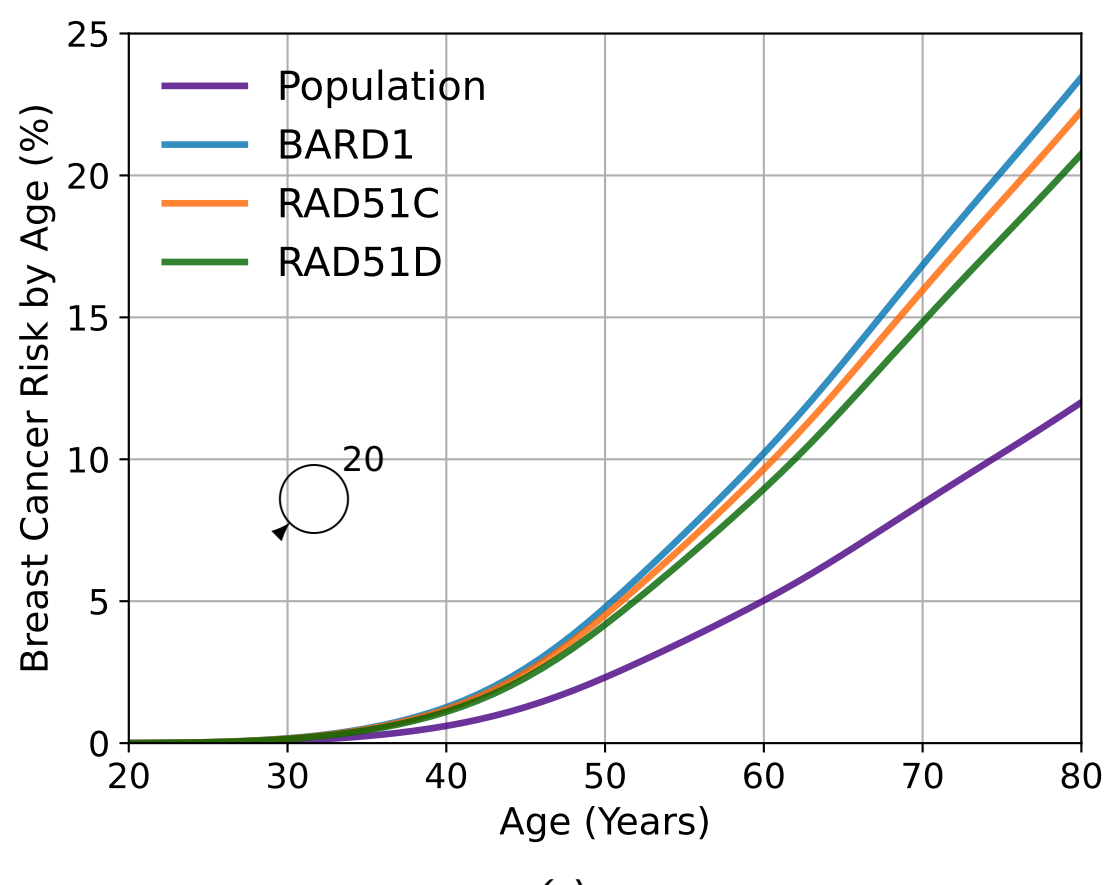
Figure 5. Predicted lifetime breast and ovarian cancer risks as a function of height for a female born in 1985 with unknown family history, comparing the updated model, where height is treated as continuous, to the previous model, where height was treated as categorical. Figures (a), (c) and (e) show breast cancer, while Figures (b), (d) and (f) show ovarian cancer risks. Figures (a) and (b) show the predicted risk as a function of height, while Figures (c) and (d) show the probability density/mass of risk as a function of height. Predictions are based on UK cancer incidences. Figures (e) and (f) show the log (base 10) of the root-mean-squared relative discretisation error as a function of the number of bins. The error was taken to be the absolute difference between the value and the asymptotic extrapolation of the measurements as a function of the number of bins. The average is taken over 100 heights that are spaced 1% apart, from 0.5% to 99.5%.

Table 1. The parameters used to include the effects of rare high- and intermediate-risk pathogenic variants in the models.

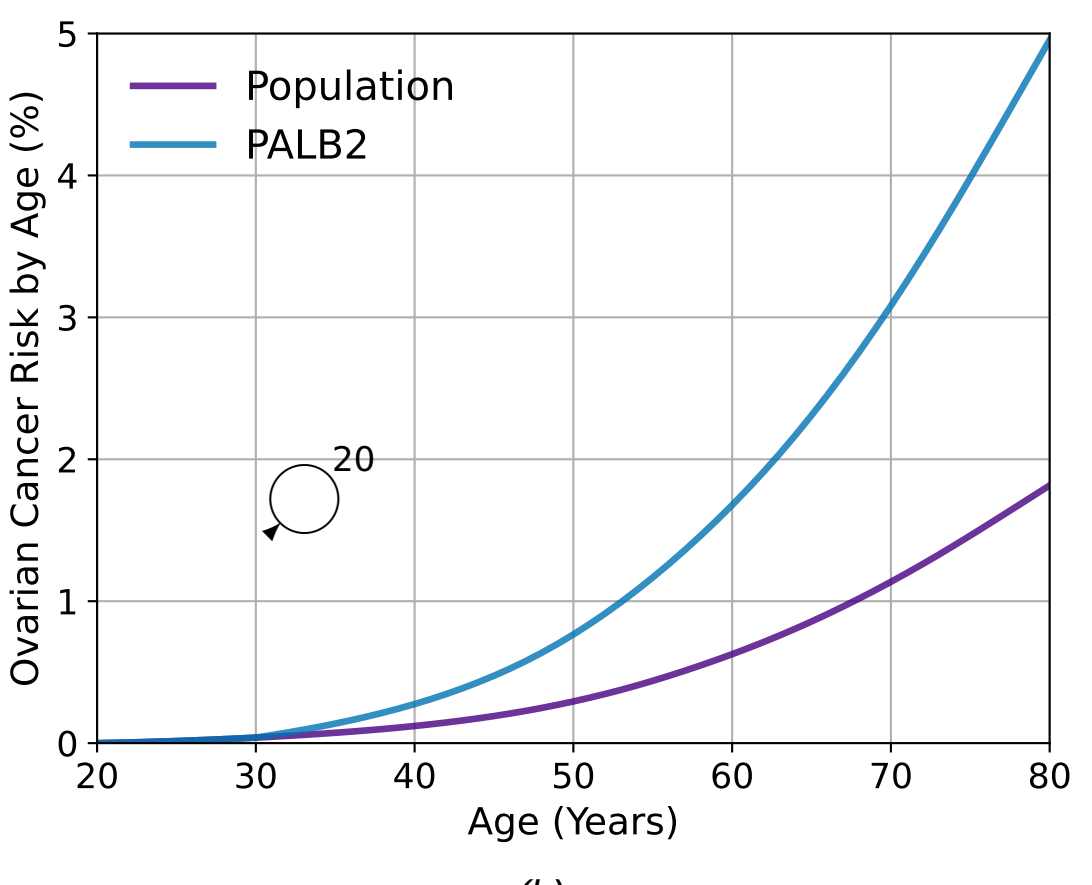
GENE	ALLELE FREQ	SS	RR OF FEMALE BREAST CANCER (95% CI)	RR OF EOC (95% CI)	RR OF MALE BREAST CANCER (95% CI)	RR OF PROSTATE CANCER	RR OF PANCREATIC CANCER (95% CI)
BRCA1	BOADICEA 0.0006394 EOC Model 0.0007947	0.89	1 $\quad age < 20$ $\exp(3.0146 + 0.02412 \times age)$ $\quad 20 \le age \le 29$ $\exp(6.0707 - 0.07775 \times age)$ $\quad 30 \le age \le 39$ $\exp(4.2511 - 0.03226 \times age)$ $\quad 40 \le age \le 49$ $\exp(4.2086 - 0.03141 \times age)$ $\quad 50 \le age \le 79$	1 $\quad age < 30$ $\exp(-3.55 + 0.1986 \times age)$ $\quad 30 \le age \le 39$ $\exp(7.1776 - 0.06959 \times age)$ $\quad 40 \le age \le 49$ $\exp(4.5236 - 0.01651 \times age)$ $\quad 50 \le age \le 79$	8	$1.82 \quad age < 65$ $0.84 \quad age \ge 65$	$3.10 \quad age < 65$ $1.54 \quad age \ge 65$
BRCA2	BOADICEA 0.00102 EOC Model 0.002576	0.96	1 $\quad age < 20$ $\exp(3.2153 - 0.008815 \times age)$ $\quad 20 \le age \le 29$ $\exp(4.28945 - 0.04462 \times age)$ $\quad 30 \le age \le 39$ $\exp(3.96865 - 0.0366 \times age)$ $\quad 40 \le age \le 49$ $\exp(1.8169 + 0.006435 \times age)$ $\quad 50 \le age \le 59$ $\exp(-0.2606 + 0.04106 \times age)$ $\quad 60 \le age \le 69$ 13.0991 $\quad 70 \le age \le 79$	1 $\quad age < 40$ $\exp(-9.708 + 0.2427 \times age)$ $\quad 40 \le age \le 53$ $\exp(6.50334 - 0.05751 \times age)$ $\quad 54 \le age \le 57$ $\exp(11.3175 - 0.140513 \times age)$ $\quad 58 \le age \le 69$ 4.4 $\quad 70 \le age \le 79$	80	$7.33 \quad age < 65$ $3.39 \quad age \ge 65$	$5.54 \quad age < 65$ $1.61 \quad age \ge 65$
PALB2	0.00064	0.92	$1 \quad age < 20$ $9.1 \quad 20 \le age \le 24$ $8.97 \quad 25 \le age \le 29$ $8.85 \quad 30 \le age \le 34$ $8.54 \quad 35 \le age \le 39$ $8.02 \quad 40 \le age \le 44$ $7.31 \quad 45 \le age \le 49$ $6.55 \quad 50 \le age \le 54$ $5.92 \quad 55 \le age \le 59$ $5.45 \quad 60 \le age \le 64$ $5.10 \quad 65 \le age \le 69$ $4.82 \quad 70 \le age \le 74$ $4.56 \quad 75 \le age \le 79$	$1 \quad age < 30$ $2.91\ (1.40 - 6.04) \quad age \ge 30$	$1 \quad age < 30$ $7.34\ (1.28 - 42.18) \quad age \ge 30$	1	$1 \quad age < 30$ $2.37\ (1.24 - 4.50) \quad age \ge 30$
CHEK2	0.00373	0.98	$1 \quad age < 20$ $\exp(1.605325 - 0.0148367 \times age) \quad age \ge 20$	1	1	1	1
ATM	0.0018	0.94	$2.10\ (1.17 - 2.57)$	1	1	1	1
BARD1	0.00043	0.89	$2.09\ (1.35 - 3.23)$	1	1	1	1
RAD51C	0.00035	0.78	$1.97\ (1.48 - 2.62)$	$1 \quad age < 30$ $\exp(-1.7974 + 0.07631 \times age) \quad 30 \le age < 60$ $\exp(9.7592 - 0.1163 \times age) \quad age \ge 60$	1	1	1
RAD51D	0.00035	0.86	$1.82\ (1.34 - 2.47)$	$1 \quad age < 30$ $\exp(-2.88662 + 0.09656 \times age) \quad 30 \le age < 58$ $\exp(5.99144 - 0.05651 \times age) \quad age \ge 58$	1	1	1
BRIP1	0.00071	0.95	1	$3.41\ (2.12 - 5.54)$	1	1	1

"Allele Freq" is the pathogenic variant allele frequency in the general population, SS is the screening sensitivity, RR is the relative risk, relative to the general population and EOC is epithelial tubo-ovarian cancer. The BOADICEA model includes the effects of *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *BARD1*, *RAD51C* and *RAD51D*, while the EOC Model includes the effects of *BRCA1*, *BRCA2*, *RAD51D*, *RAD51C*, *BRIP1* and *PALB2*. The updated parameters are the allele frequencies for *PALB2*, *CHEK2*, *ATM*, *RAD51C*, *RAD51D*, *BRAD1* and *BRIP1*[15], the SS for pathogenic variants in for all

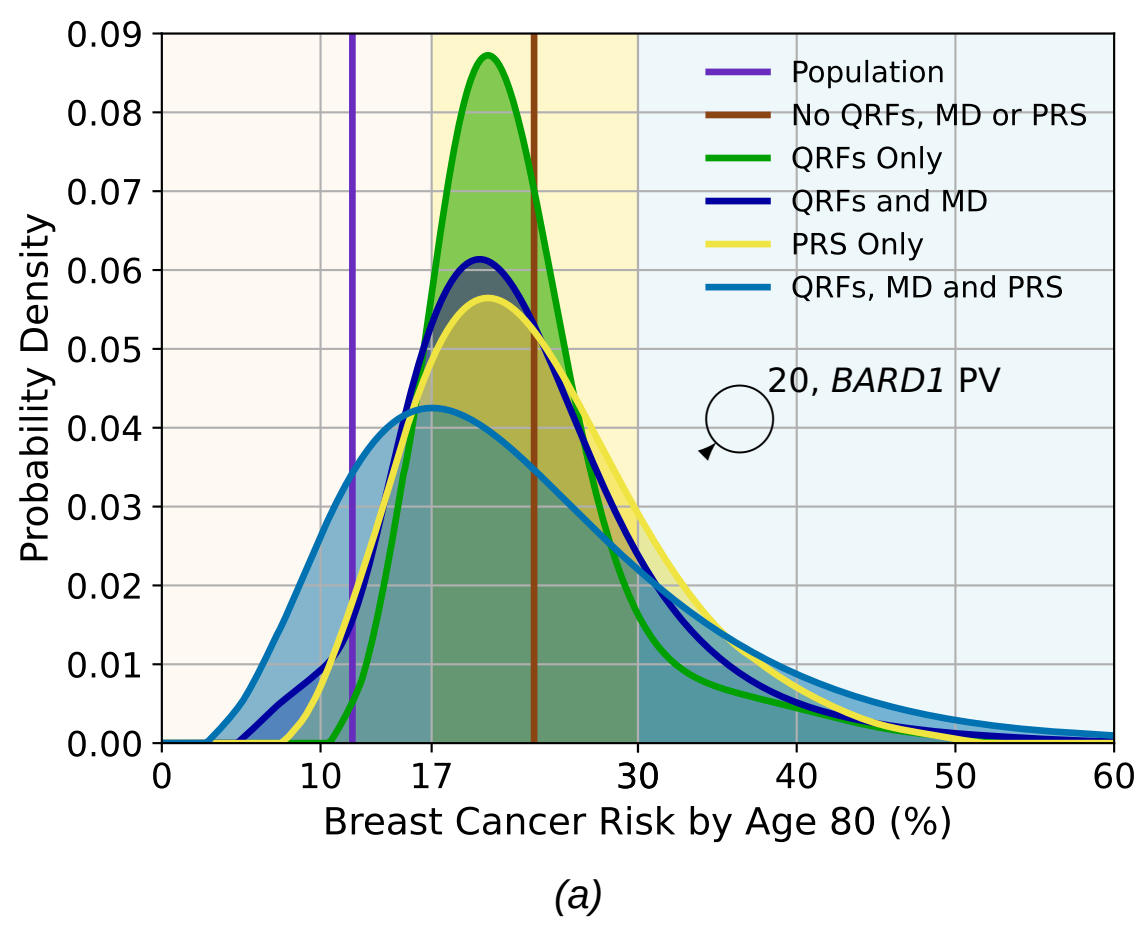
genes [15], the RR for female breast cancer for *ATM*, *BARD1*, *RAD51C* and *RAD51D* [15 24], and the EOC, male breast cancer and the pancreatic cancer RRs for *PALB2*[17]. All other parameters are as previously published[8-11 21].


Population
BARD1
RAD51C
RAD51D
20
Breast Cancer Risk by Age (%)
Age (Years)
0
5
10
15
20
25
20
30
40
50
60
70
80

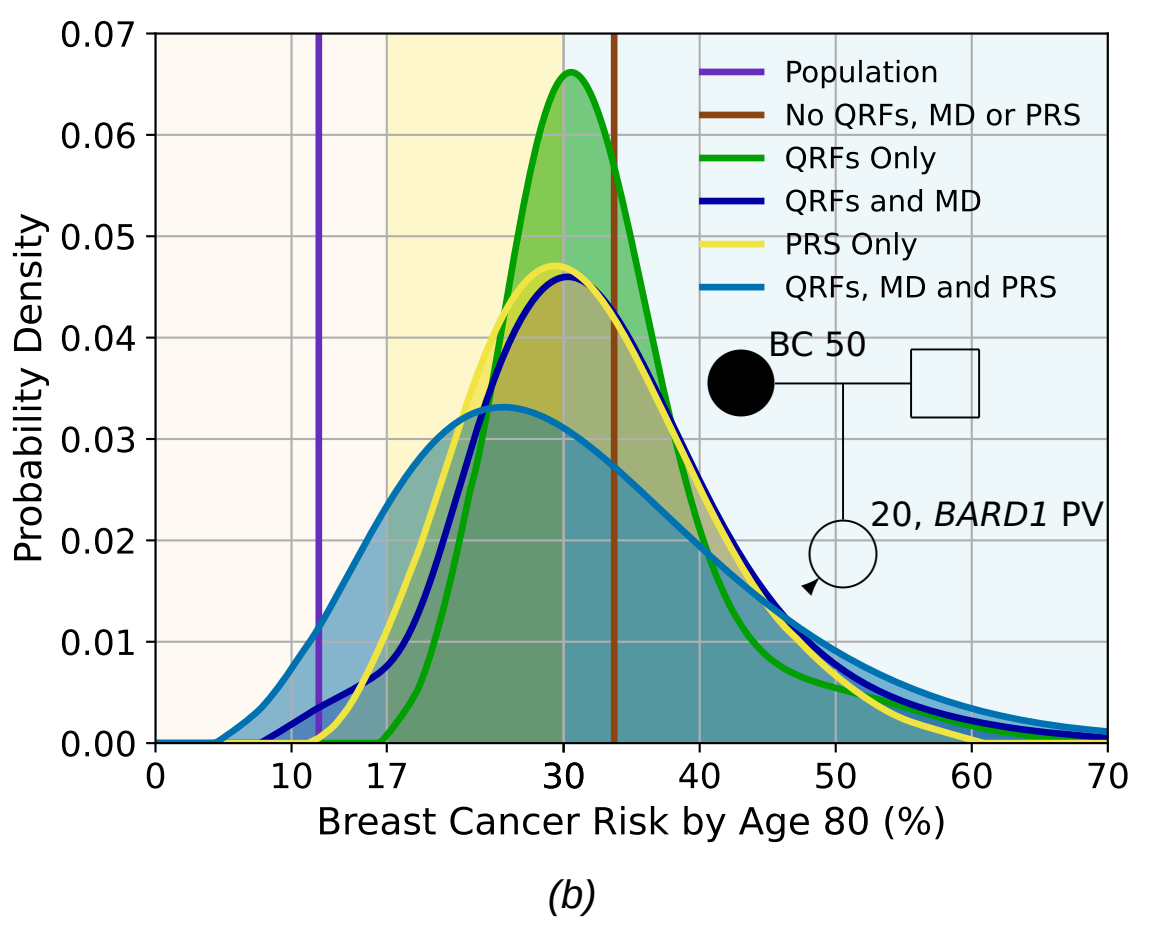
(a)


Population
PALB2
20
Ovarian Cancer Risk by Age (%)
Age (Years)
0
1
2
3
4
5
20
30
40
50
60
70
80

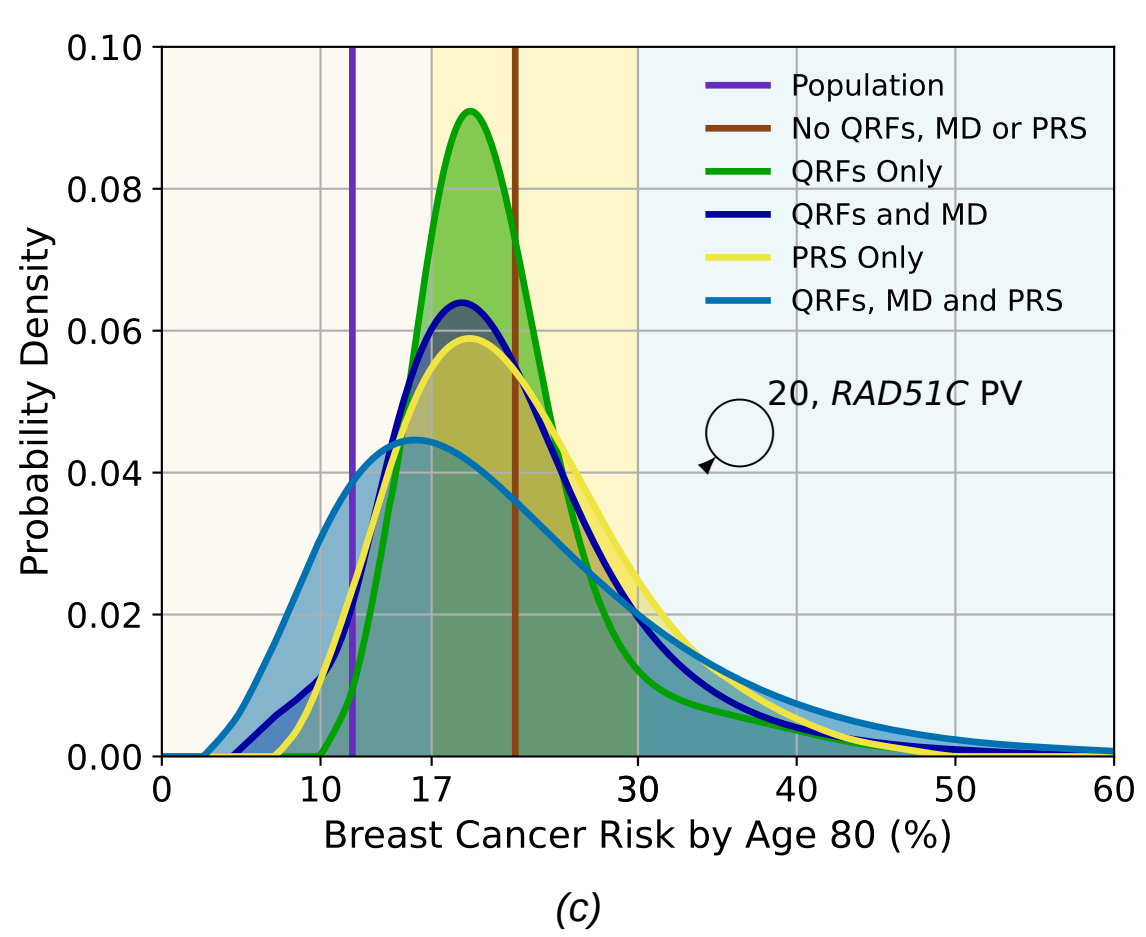
(b)


Probability Density
Breast Cancer Risk by Age 80 (%)
Population
No QRFs, MD or PRS
QRFs Only
QRFs and MD
PRS Only
QRFs, MD and PRS
20, *BARD1* PV

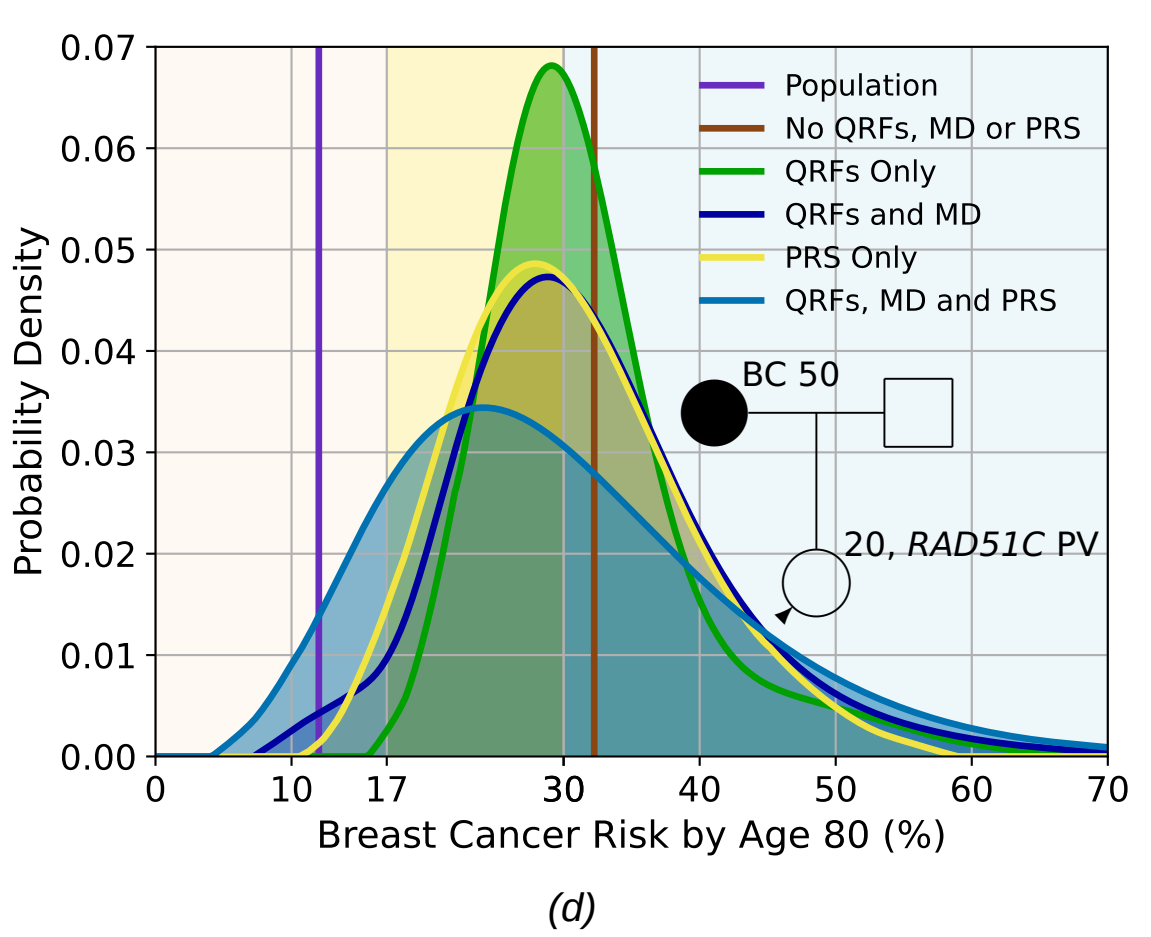
(a)


Probability Density
Breast Cancer Risk by Age 80 (%)
Population
No QRFs, MD or PRS
QRFs Only
QRFs and MD
PRS Only
QRFs, MD and PRS
BC 50
20, *BARD1* PV

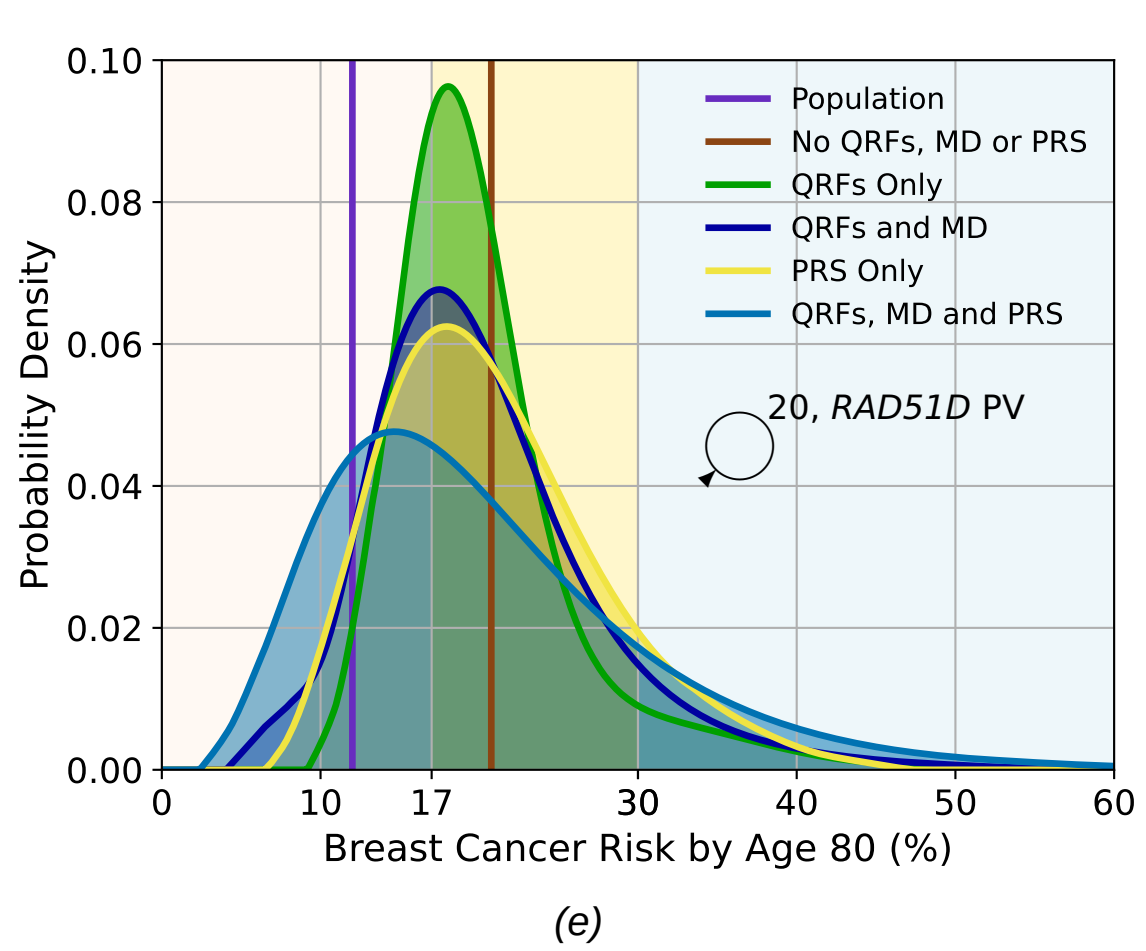
(b)


Probability Density
Breast Cancer Risk by Age 80 (%)
Population
No QRFs, MD or PRS
QRFs Only
QRFs and MD
PRS Only
QRFs, MD and PRS
20, *RAD51C* PV

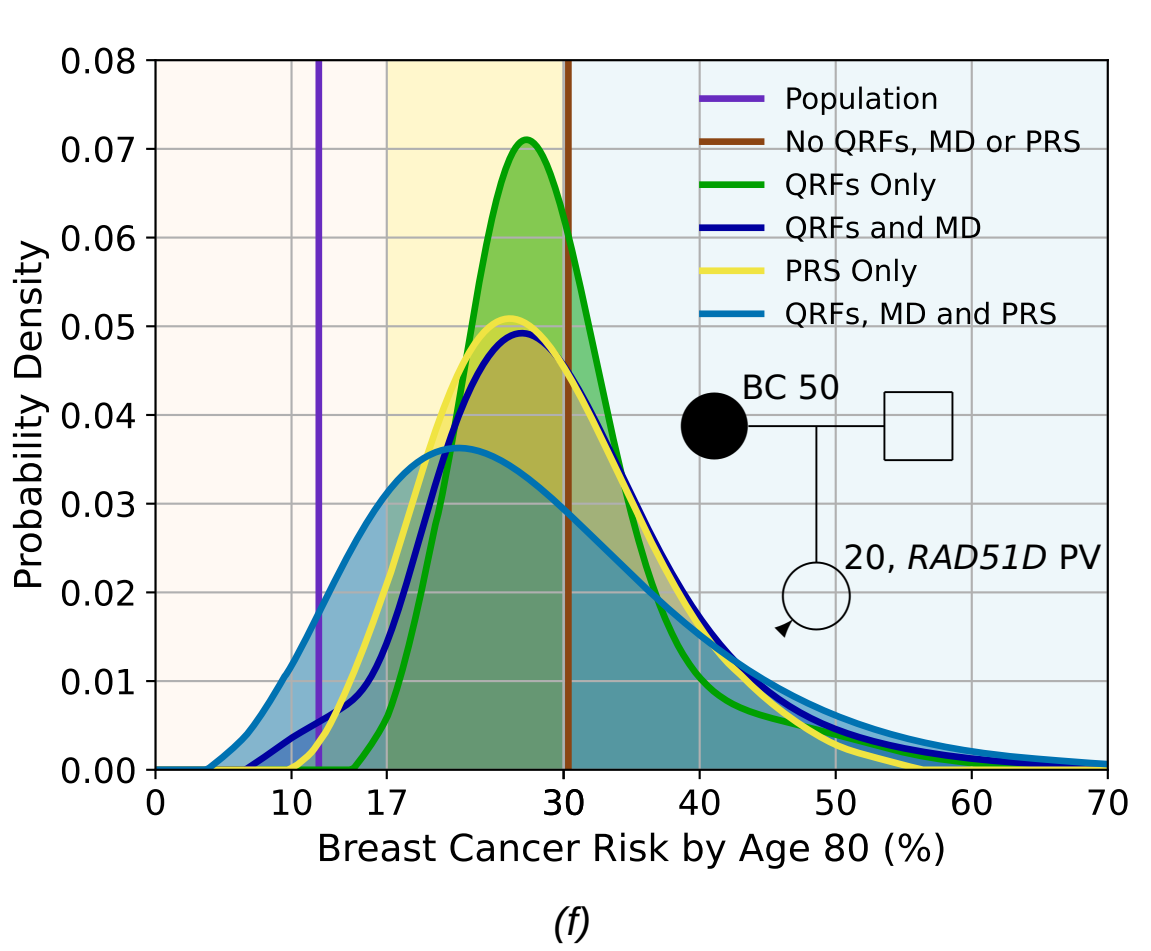
(c)


Probability Density
Breast Cancer Risk by Age 80 (%)
Population
No QRFs, MD or PRS
QRFs Only
QRFs and MD
PRS Only
QRFs, MD and PRS
BC 50
20, *RAD51C* PV

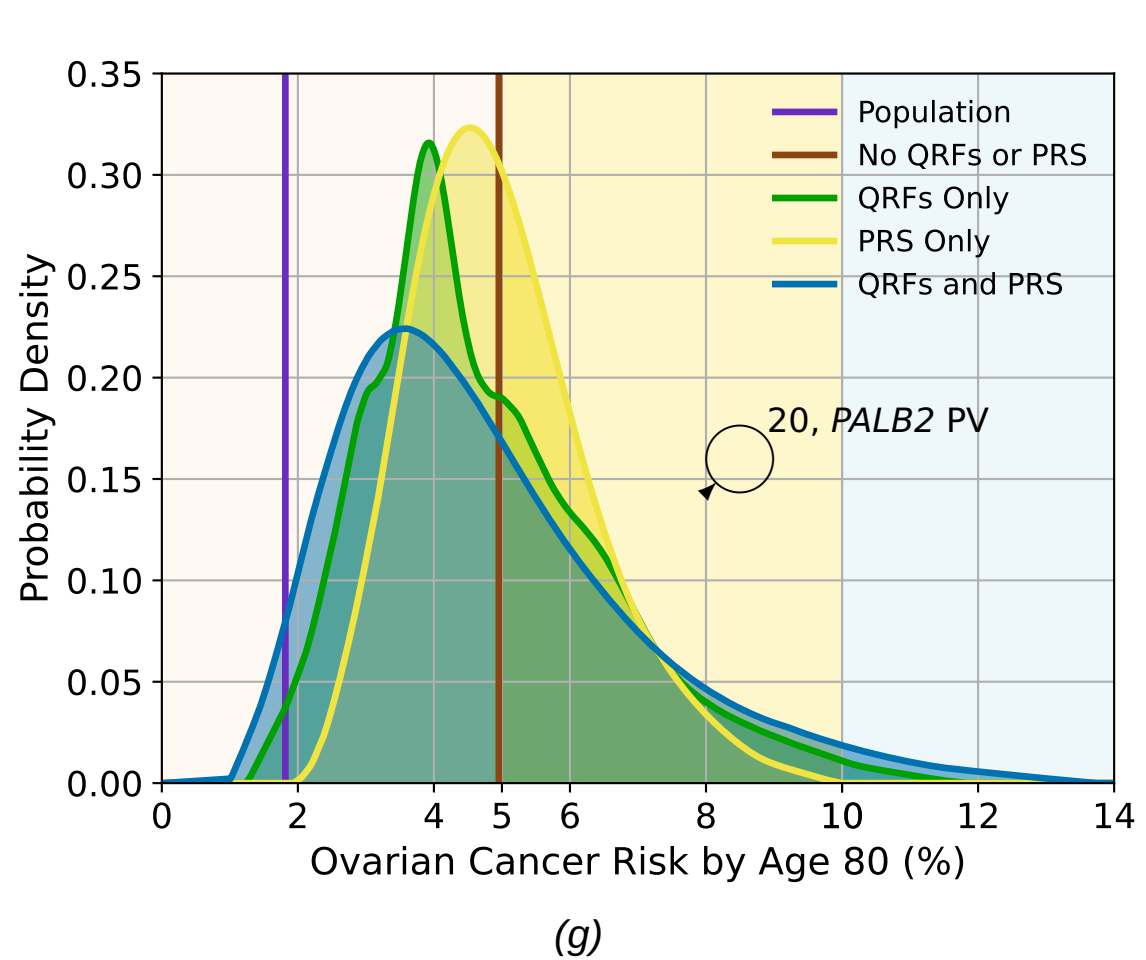
(d)


Probability Density
Breast Cancer Risk by Age 80 (%)
Population
No QRFs, MD or PRS
QRFs Only
QRFs and MD
PRS Only
QRFs, MD and PRS
20, *RAD51D* PV

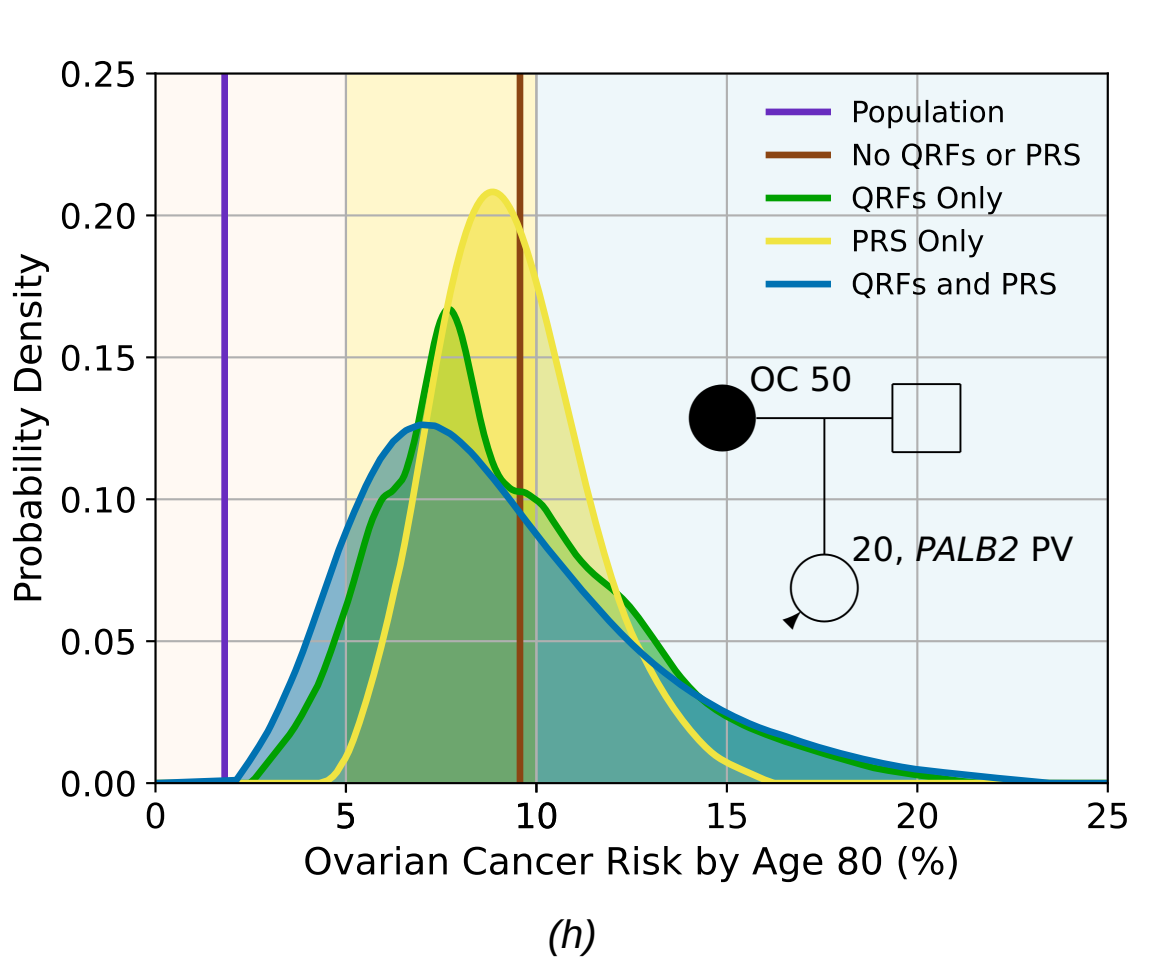
(e)


Probability Density
Breast Cancer Risk by Age 80 (%)
Population
No QRFs, MD or PRS
QRFs Only
QRFs and MD
PRS Only
QRFs, MD and PRS
BC 50
20, *RAD51D* PV

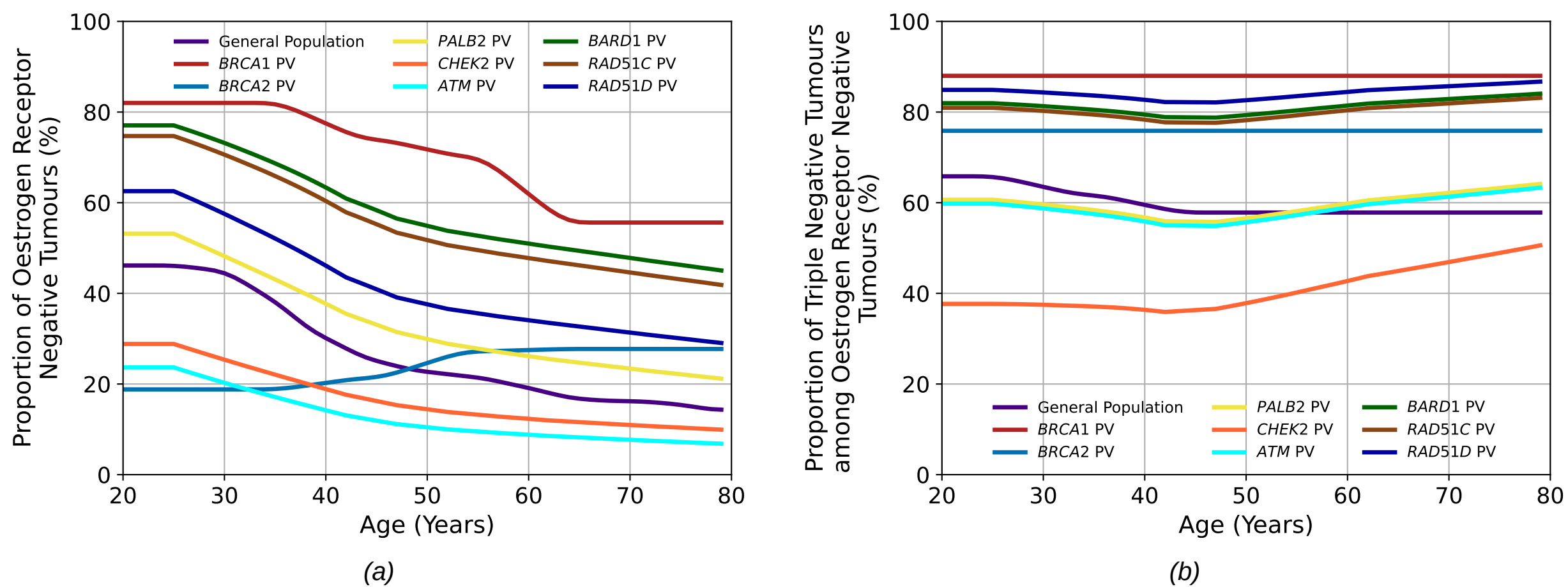
(f)

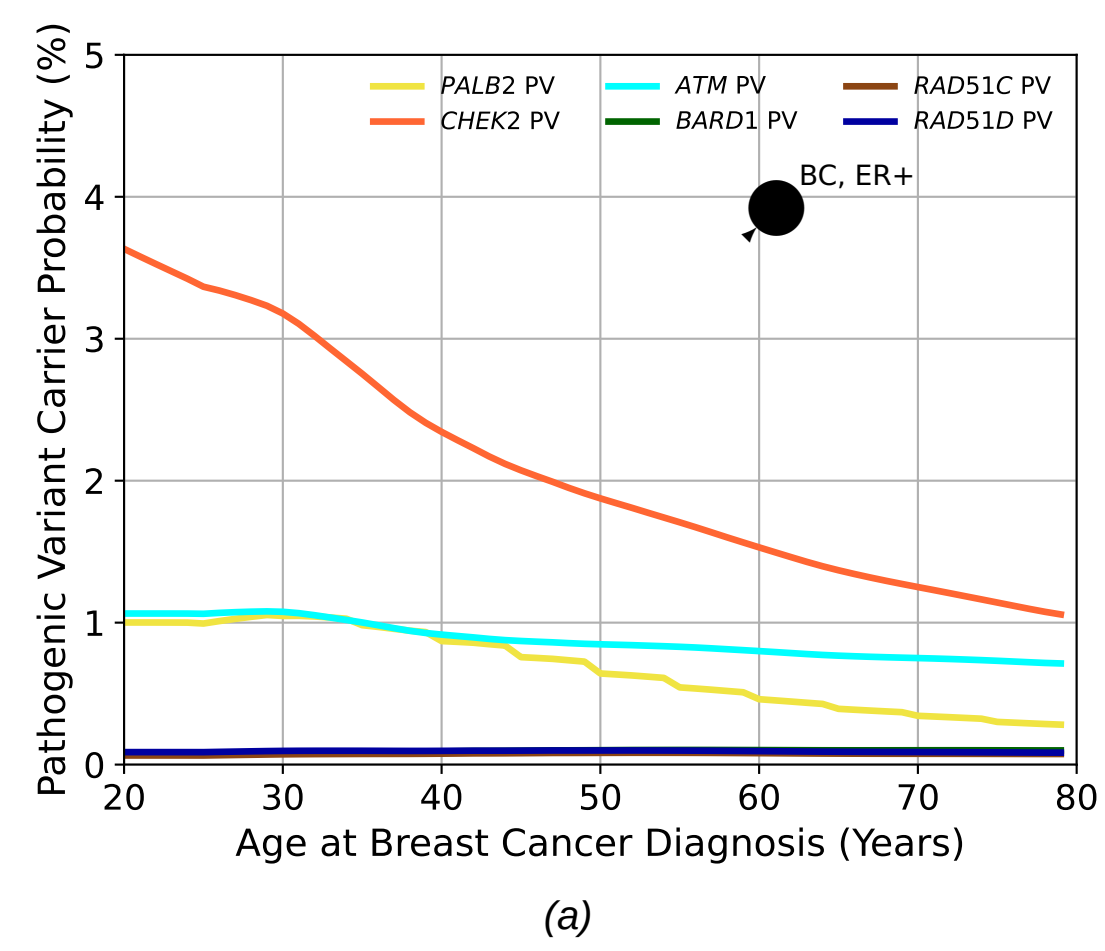

Probability Density
Ovarian Cancer Risk by Age 80 (%)
Population
No QRFs or PRS
QRFs Only
PRS Only
QRFs and PRS
20, *PALB2* PV

(g)

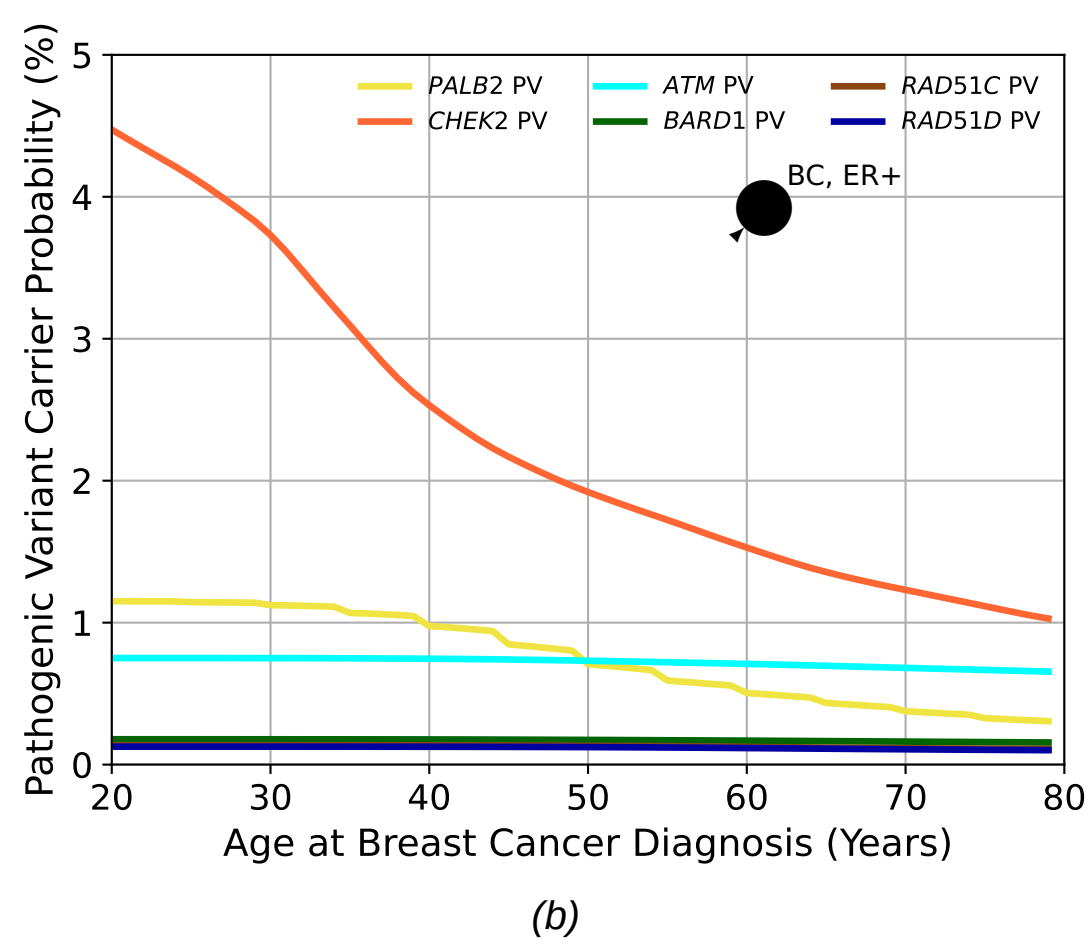

Probability Density
Ovarian Cancer Risk by Age 80 (%)
Population
No QRFs or PRS
QRFs Only
PRS Only
QRFs and PRS
OC 50
20, *PALB2* PV

(h)

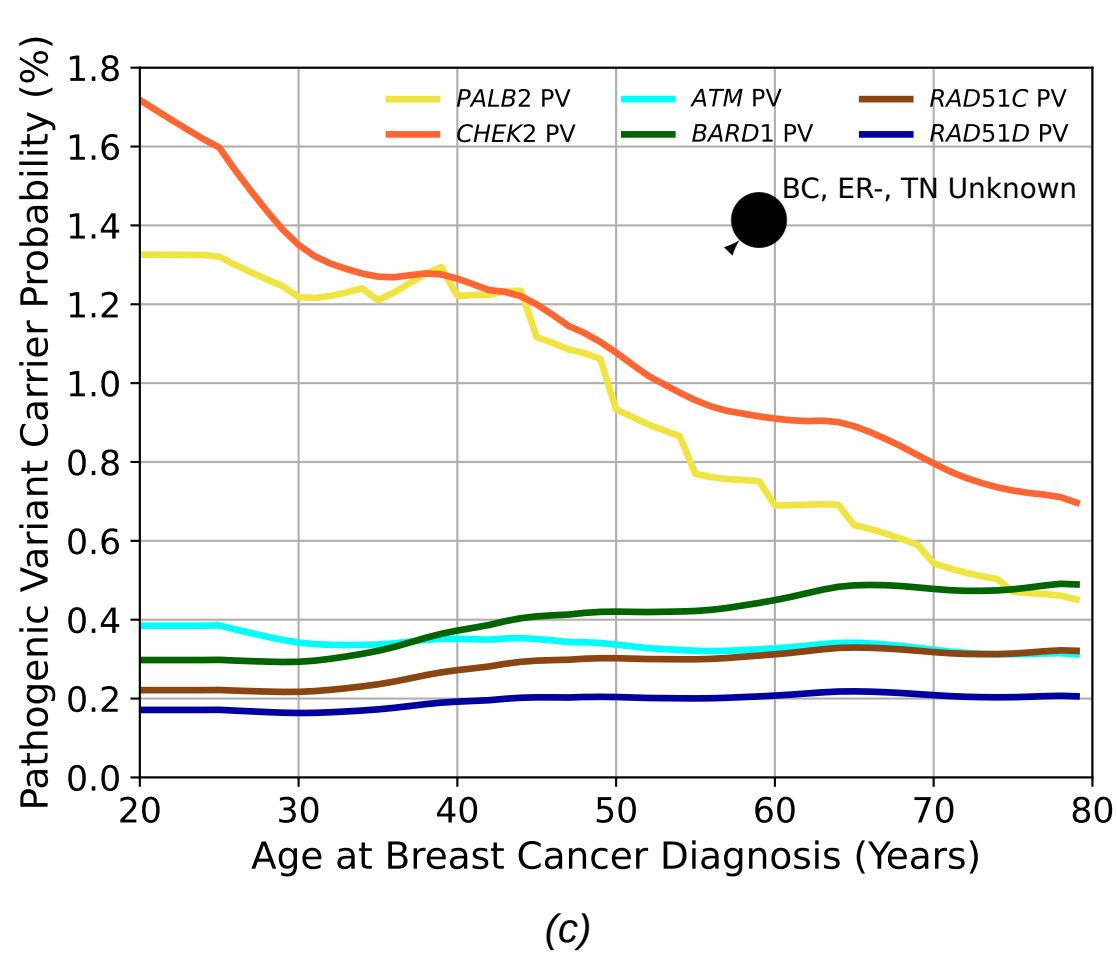

Proportion of Oestrogen Receptor Negative Tumours (%)
Age (Years)
General Population
BRCA1 PV
BRCA2 PV
PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
(a)
Proportion of Triple Negative Tumours among Oestrogen Receptor Negative Tumours (%)
Age (Years)
General Population
BRCA1 PV
BRCA2 PV
PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
(b)


PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
BC, ER+
Pathogenic Variant Carrier Probability (%)
Age at Breast Cancer Diagnosis (Years)

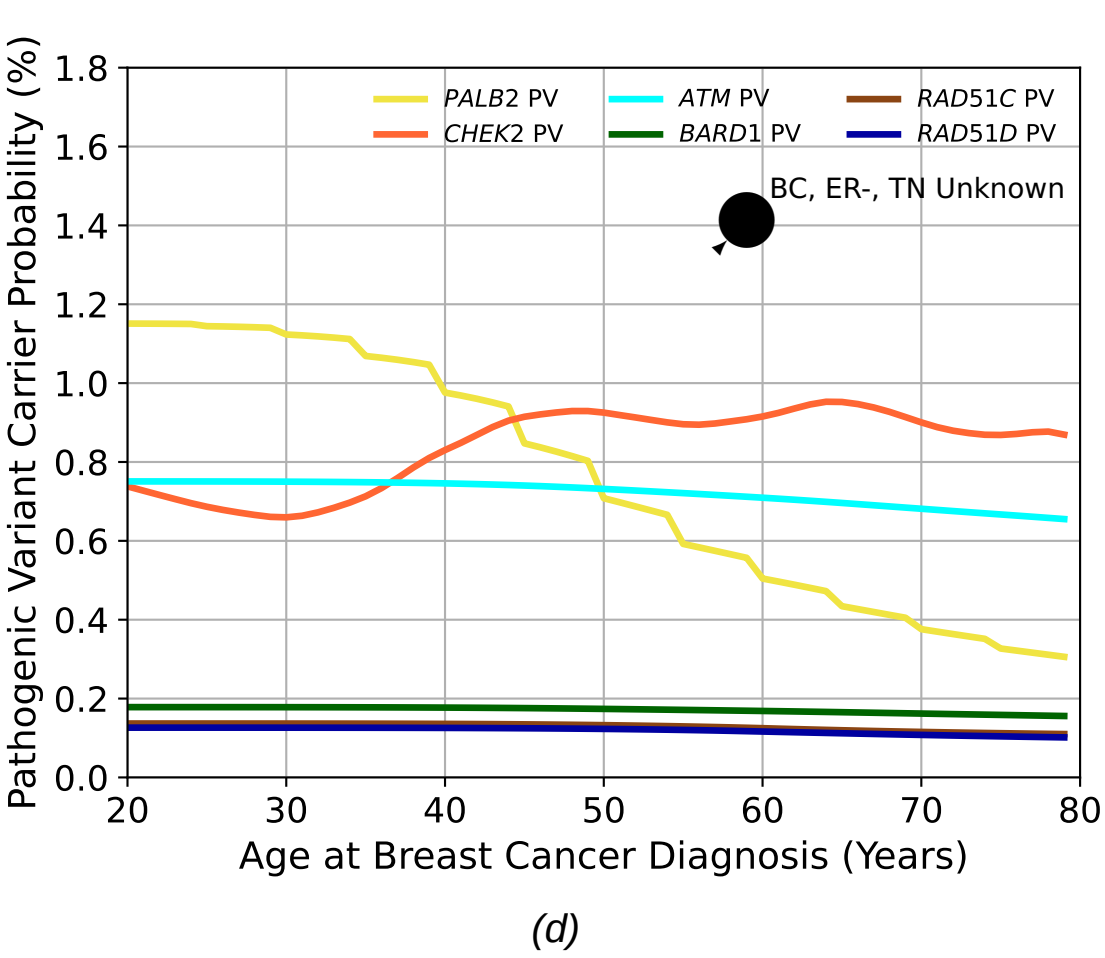
(a)


PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
BC, ER+
Pathogenic Variant Carrier Probability (%)
Age at Breast Cancer Diagnosis (Years)

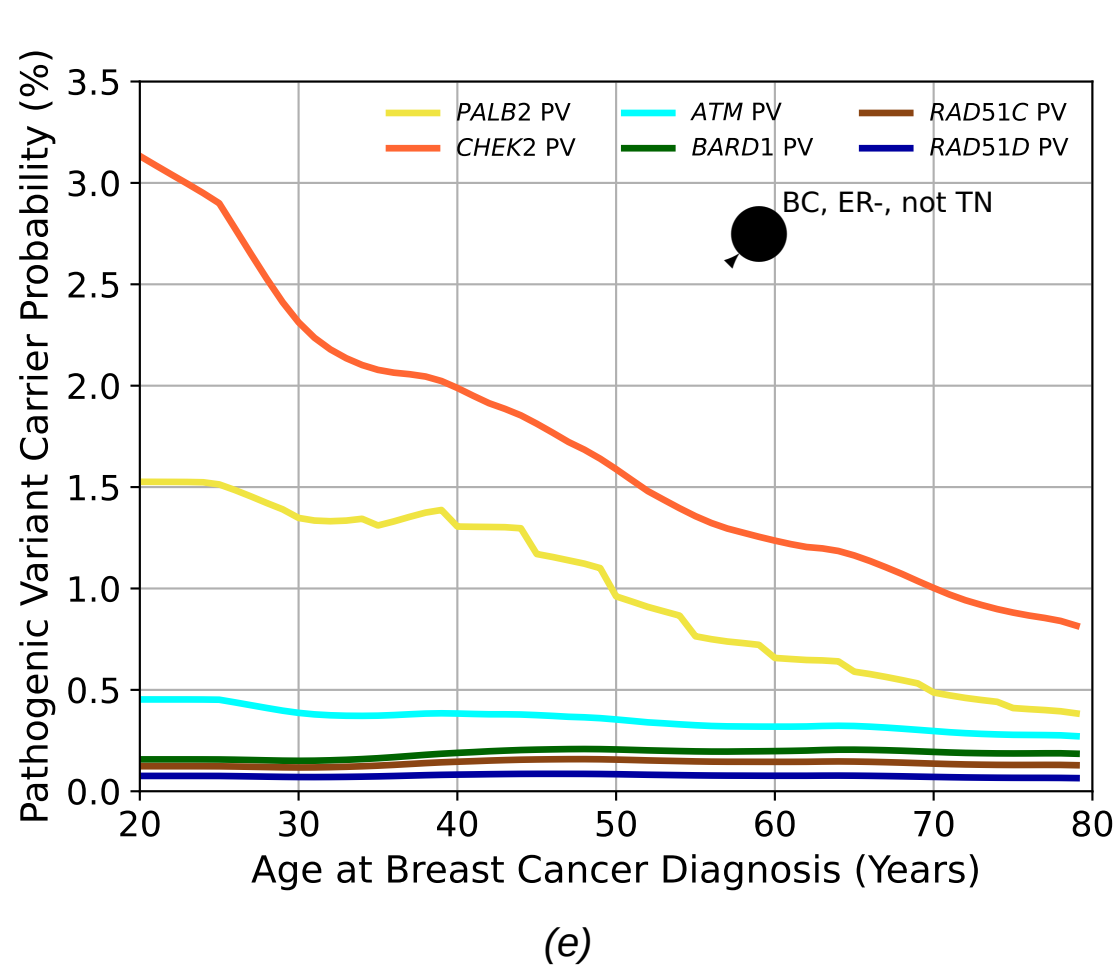
(b)


PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
BC, ER-, TN Unknown
Pathogenic Variant Carrier Probability (%)
Age at Breast Cancer Diagnosis (Years)

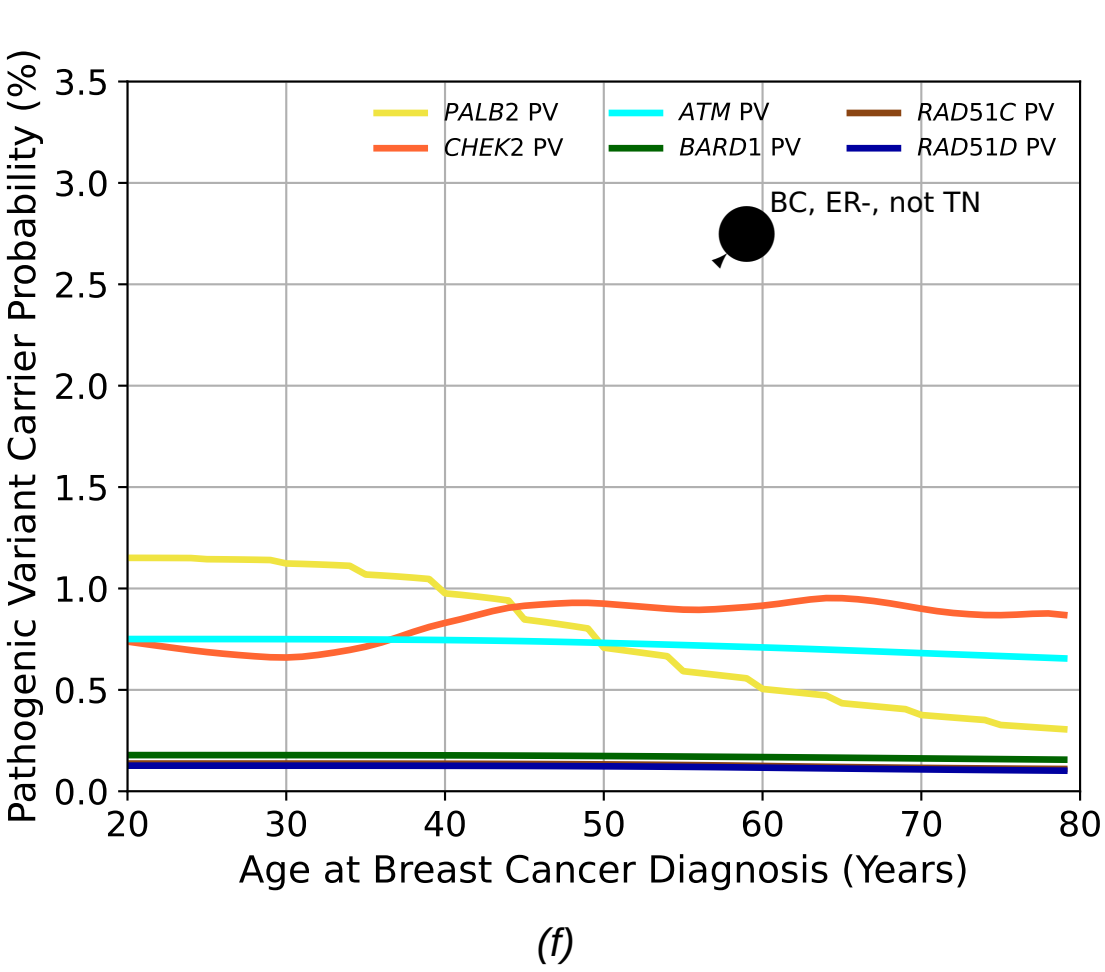
(c)


PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
BC, ER-, TN Unknown
Pathogenic Variant Carrier Probability (%)
Age at Breast Cancer Diagnosis (Years)

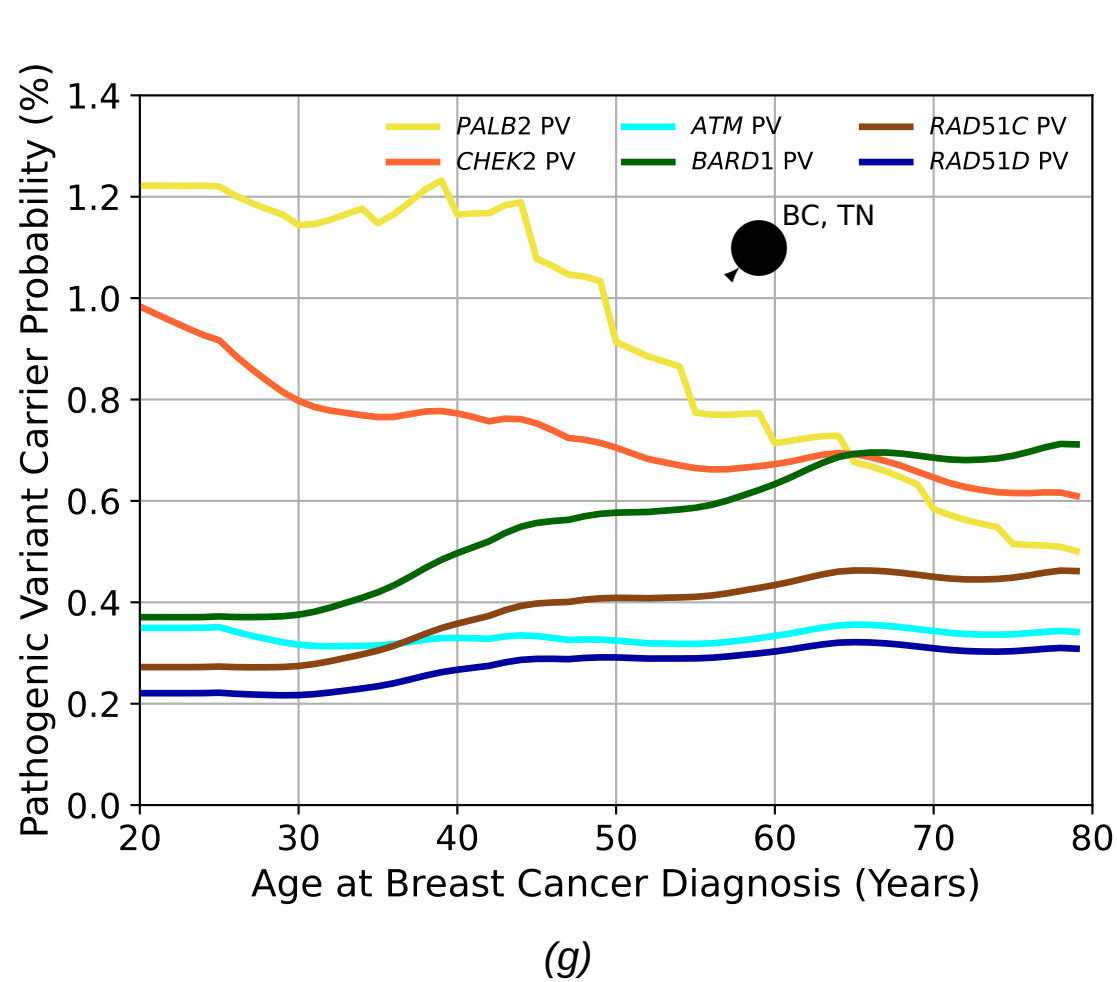
(d)


PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
BC, ER-, not TN
Pathogenic Variant Carrier Probability (%)
Age at Breast Cancer Diagnosis (Years)

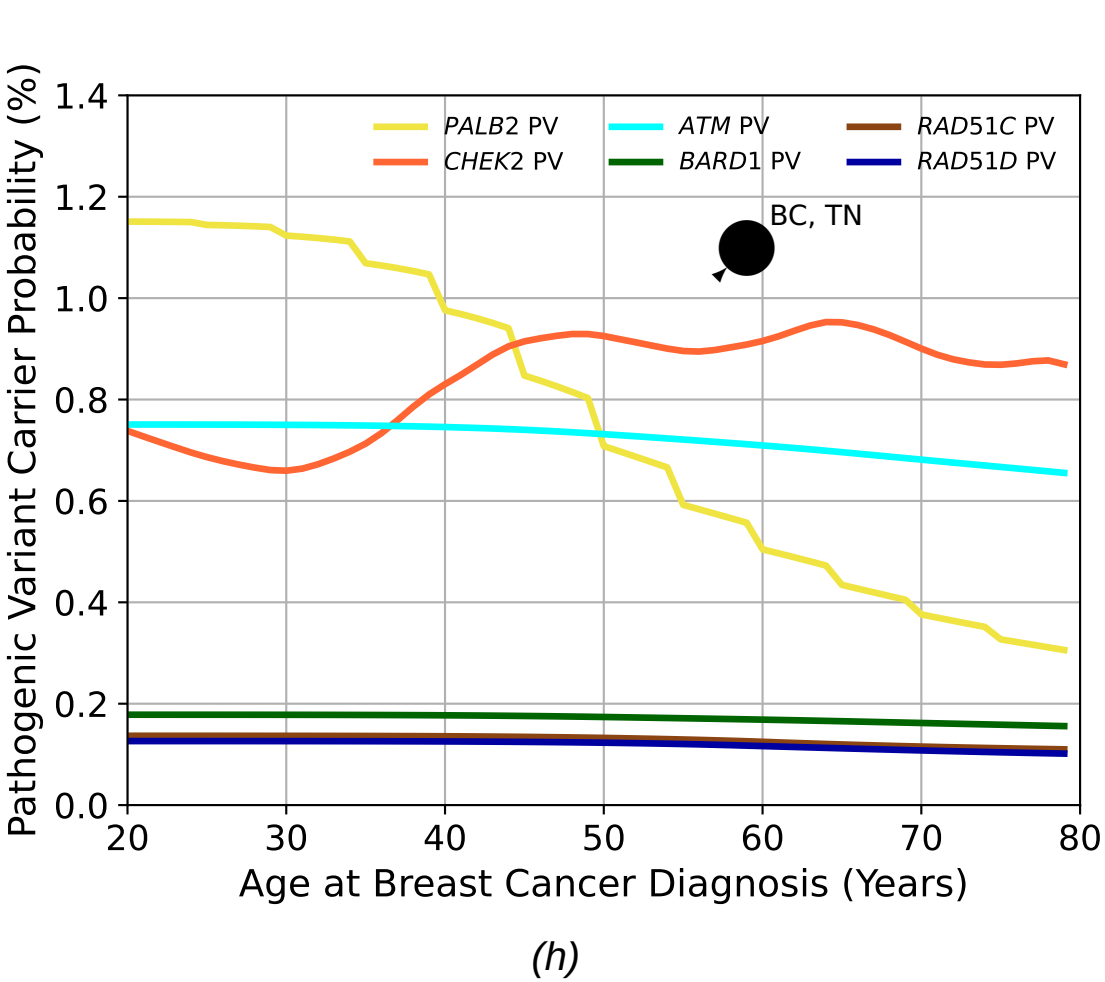
(e)


PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
BC, ER-, not TN
Pathogenic Variant Carrier Probability (%)
Age at Breast Cancer Diagnosis (Years)

(f)


PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
BC, TN
Pathogenic Variant Carrier Probability (%)
Age at Breast Cancer Diagnosis (Years)

(g)


PALB2 PV
CHEK2 PV
ATM PV
BARD1 PV
RAD51C PV
RAD51D PV
BC, TN
Pathogenic Variant Carrier Probability (%)
Age at Breast Cancer Diagnosis (Years)

(h)

Supplementary Material

Pathogenic Variants in Cancer Susceptibility Genes

Model definition

The BOADICEA and epithelial tubo-ovarian cancer (EOC) models assume that the cancer incidences for individual i at age t, $\lambda^{(i)}(t)$, depend on their underlying genotype through a model of the form:

$$\lambda^{(i)}(t) = \lambda_0(t)\exp\left(\sum_{\mu=1}^{N_{MG}+1}\left[\left(\beta_{MG\mu}(t) + \sum_{\rho}\boldsymbol{\beta}_{RF\rho\mu}(t)\cdot\boldsymbol{z}_{RF\rho}^{(i)}\right)\prod_{\nu=1}^{\mu-1}\left[\left(1-G_{\nu}^{(i)}\right)G_{\mu}^{(i)}\right]\right] + \beta_{PG}(t)\,x_P^{(i)}\right), \quad (1)$$

where $\lambda_0(t)$ is the baseline incidence (applicable to a non-PV carrier with a zero polygenotype and unknown RFs). N_{MG} is the number of major genes present in the model, which for the previous versions of both models was five. $G_{\mu}^{(i)}$ are indicator variables for the presence/absence of a PV in a major gene in person i, taking values 1 if a PV is present and 0 otherwise with $\mu = 1, \ldots, N_{MG}$ representing the genes present in the model in the dominance order and $\mu = N_{MG}+1$ corresponding to non-carriers of PVs, where $G_{N_{MG}+1}^{(i)} = 1$ for non-carriers of any PV and 0 otherwise. The cancer incidences associated with homozygous and heterozygous carriers of PVs in each gene are assumed to be the same, and the risk to carriers of PVs in more than one gene is assumed to be that of the higher-ranked PV in the dominance order. Because PVs are rare, this model can be well approximated by assuming a single locus with $N_{MG}+1$ alleles, one representing the presence of a PV in each of the N_{MG} genes and an additional wild-type allele representing absence of PVs in all genes[1]. $\beta_{MG\mu}(t)$ represent the age-specific log-relative risks (log-RRs) associated with the major genes relative to the baseline incidence. The relative risks (RR) assumed for the major genes are summarised in Table 1. $x_P^{(i)}$ is the polygenotype for individual i, assumed normally distributed in the general population with mean 0 and standard deviation 1, and $\beta_{PG}(t)$ is the age-specific log-RR per standard deviation associated with the polygene, relative to the baseline incidence[2 3]. When a PRS is known, the polygenotype is decomposed into an observed and residual component where the observed component is given by the PRS[4]. ρ indexes the RFs that are present in the model, which are modelled as categorical factors. $\boldsymbol{\beta}_{RF\rho\mu}(t)$ is the vector (length $\kappa_\rho - 1$ were κ_ρ is the number of categories for RF ρ, with one category being the baseline) of age-specific log-RRs associated with RF ρ, which may depend on the major genotype μ, and $\boldsymbol{z}_{RF\rho}^{(i)}$ is the corresponding vector of indicator variables (0 or 1) that indicate the category of RF ρ for individual i (1 for the observed category, 0 otherwise, with all elements 0 for the baseline). The baseline incidences $\lambda_0(t)$ are determined so that the total age-specific incidences, summed over the RFs and genotypes, agree with the population incidence (given the assumed population distributions and RRs)[2 5]. The population incidences are birth-cohort and country-specific, but this dependence is omitted from equation (1) for clarity of notation. The RRs and distributions of the RF have been described elsewhere[4 6]. To allow appropriately for missing

RF information, only those RFs measured on a given individual are considered (thus, the baseline incidence, $\lambda_0(t)$ are determined for each individual dependent on their measured RFs).

The models assume that RRs associated with PVs in the major genes are log-additive (multiplicative) with the RFs and the polygenic component. The model also assumes that the PVs and the PRS combine multiplicatively (conditional on other factors).

The models evaluate pedigree likelihoods using the MENDEL software[7]. As MENDEL considers only finite discrete genotypes, the polygenotype is approximated by the hypergeometric polygenic model [1 5 8].

Both models consider family history of breast cancer (BC), EOC, pancreatic cancer (PaC) and prostate cancer (PrC). The incidences of each cancer are assumed independent, conditional on the genotypes and RFs in the model. In BOADICEA, EOC, PaC and PrC are assumed to depend only on the major genotype. Correspondingly, in the EOC model, BC, PaC and PrC are assumed to depend only on the major genotype.

Adjusting the residual polygenic component after the inclusion of new major genes

The variance due to PVs in each gene at age t is given by:

$$var(t,\mu) = \log\left(\frac{(1-f_\mu)^2 + f_\mu(2-f_\mu)\exp(2\,\beta_{MG\mu}(t))}{\left((1-f_\mu)^2 + f_\mu(2-f_\mu)\exp(\beta_{MG\mu}(t))\right)^2}\right),$$

where f_μ is the population allele frequency of gene μ; the variance components are assumed to be additive. This process also considered the updated RR and PV frequencies for the previously included genes. For BOADICEA, the overall BC polygenic variance was $4.83 - 0.5961 \times t$ for females and 1.4 for males, while for the EOC model, the overall EOC polygenic variance was 1.434 [2 3].

Allele Frequencies

Allele frequencies for all genes, except *BRCA1* and *BRCA2,* were taken from the BRIDGES study [9]. The frequencies were based on the frequency of protein-truncating variants in European ancestry controls. To account for the incomplete sensitivity of the sequencing as performed in BRIDGES, the frequencies were adjusted by dividing by $cs(1-v)$, where c is the proportion of the coding sequence of each gene determined to be callable, s is the proportion of variants in the called sequence across all genes that were detected (estimated to be 0.957), and v is the proportion of the pathogenic variants expected to be copy variants. For *CHEK2*, the adjustment was applied to variants excluding c.1100delC. Details are given in the Supplementary Material of Dorling *et al.*[9]. For *BRIP1, v* was assumed to be 0.05. The *BRCA1* and *BRCA2* frequencies from the previous versions of BOADICEA and the EOC model were used for consistency.

Sensitivities

The default sensitivities are based on the assumption that protein truncating variants and known pathogenic missense variants are detected with close to 100% sensitivity in clinical tests but that, except for *BRCA1* and *BRCA2*, large rearrangements are not detected. The

sensitivities are therefore given by $1 - v$, as above. For *BRCA1* and *BRCA2*, sensitivities were defined by assuming that the main source of insensitivity was missense variants not classified as pathogenic – the frequencies of these variants have been estimated by Dorling et al[10]. Were large re-arrangements not tested for, the corresponding sensitivities for *BRCA1* and *BRCA2* would be reduced to ~76% and 95% respectively (owing the much higher frequency of large re-arrangements in *BRCA1*).

BRCA2: ovarian cancer relative risks updates

Previous estimates of the EOC relative risks for *BRCA2* PV carriers were obtained during the BOADICEA model fitting process, using complex segregation analysis in families with *BRCA2* PVs[2]. This involved fitting models in which the log-relative risks were piecewise linear functions of age. Due to the very small number of EOCs diagnosed in ages 65 years and over in the original dataset, the RR was estimated to decrease rapidly from 23.7 at age 58, to 1.59 at age 69 and remain constant at that level thereafter. However, more recent data suggest that the EOC RRs for ages 70 and over are higher[11]. The original RR estimate of 1.59 may result in an underestimation of risks for older *BRCA2* carriers. We therefore updated the log-RR function included in the model by re-deriving the piecewise log-RR linear function such that the EOC RR decreases less rapidly from 23.7 at age 58 to 4.4 for ages 70 and over. The RR=4.4 estimate used for ages 70 and over was obtained from a prospective cohort analysis of *BRCA2* PV carriers[11].

The updated log-RR EOC parameters for ages 58 and over for *BRCA2* carriers are shown in Table 1 and the resulting age-specific EOC cumulative risks are shown in Figure s3.

Population Incidences

The BOADICEA and EOC models both allow population customisation via population-specific incidences[4 6 12]. Here the models are extended with incidences from the Netherlands, France, Slovenia and Estonia. Incidences for the Netherlands were taken from Statistics Netherland for 1950-1988 and the Netherlands Cancer Registry for 1989-2017, where BC incidences exclude ductal carcinomas in situ, as these are not included in the models [13 14]. Incidences for France were taken from CI5Plus and CI5 for 1977-1989 using nine registries and from INCa/Santé Public France for 1990-2018 [15-17]. Incidences for Slovenia covering 1961-2016 were taken from the Slovenian Cancer Registry [18]. Incidences for Estonia covering 1968-2018 were taken from the Estonian National Institute for Health Development [19]. Predicted lifetime breast and EOC risks using these incidences are shown in Figure s1.

Incidences for some of the existing regions were updated using data from more recent calendar years. For the UK, incidences covering 2011-2017 were added [20]. For Denmark, Finland, Iceland, Norway and Sweden, incidences covering 2011-2018 were added [21 22]. For Australia, incidences covering 2011-2017 were added [23]. For the USA, incidences covering 2013-2018 from 21 registries were added [24]. For New Zealand, incidences covering 2010-2018 were added [25 26]. For Canada, incidences covering 2011-2018 were added [27]. Figure s2 (a) shows the updated incidences' effects on the cohort incidences for UK female breast cancer incidences for those born in the 1980s.

The models use calendar-specific population incidences to calculate cohort-specific incidences [2], where the cohorts are defined by decadal birth year ranges (1910-1919, 1920-1929, 1930-1939, 1940-1949, 1950-1959, 1960-1969, 1970-1979 and 1980-1989 with individuals born before/after the first/last cohort, assumed to have the same incidences as the first/last cohort). The original model used UK incidences from CI5, which reported calendar incidences averaged in 5-year calendar-period bins [2 15]. Cohort incidences were then taken as those for someone born in the middle year of each range to represent that cohort (1915 for 1910-1919 etc.). However, some of the other regions have smaller populations and report annual-calendar-period specific incidences. For these populations, especially for cancers with low incidences (e.g., EOC and male BC), using a single year to represent the cohort can lead to cohort incidences dominated by year-on-year calendar fluctuations. The methodology was refined by deriving new sets of cohort incidences. In these, the age-specific incidences for an individual in the cohort were taken as the average of the age-specific incidences applicable to those born in each year of the birth-cohort range. The average age- and cohort-specific incidences were then smoothed using LOWESS with linear regression and a bandwidth of 0.2. Figure s2 (b) shows the effects of the new averaging method on cohort incidences for Estonian male breast cancer incidences for those born in the 1920s.

Further, previously, incidences for years before/after the earliest/latest calendar year were taken to be the same as those in the earliest/latest calendar year available. Again, for regions with small populations presenting annual calendar-period incidences and cancers with low incidences, the cohort incidences can be adversely affected by statistical anomalies present in incidences of the earliest/latest calendar year. The methodology was refined with incidence for years before/after the earliest/latest calendar year taken as the average of the first/last five years of the available annual calendar-period incidences.

Algorithm optimisation

The BOADICEA future risk calculations rely on calculating pedigree likelihoods under the assumed genetic models of inheritance[2]. The inclusion of additional genes (*RAD51C, RAD51D, BARD1)* in the model resulted in a substantial increase in runtime. This is further compounded by the fact that separate pedigree likelihood calculations are required for risk predictions at multiple future time-points (e.g. in annual, or 5-year intervals). To reduce the programme runtime we re-formulated underlying algorithm to calculate the future risks as follows.

BOADICEA calculates the probability that an individual develops breast (or ovarian) cancer over a given time period, given the age of the proband, the genotypes, other risk factors, and family history:

$$P(D(t_1)|D(t_0), D_R, \underline{z}, \underline{\theta})$$

Where $D(t)$ is the phenotype of the proband at time t, D_R represents the phenotypes of the all the relatives, $\underline{z}$ are the risk factors measured on the proband and $\underline{\theta}$ are the genetic model parameters (allele frequencies, relative risks etc). t_0 is the current age of the proband and t_1

the future age at which the predictions are being made. In practice, these are calculated as the ratio of two pedigree likelihoods[2]:

$$\frac{P(D(t_1), D_R, \underline{z}, \underline{\theta})}{P(D(t_0), D_R, \underline{z}, \underline{\theta})}$$

The numerator and denominator probabilities, are the probabilities of the full set of phenotypes in the pedigree at times t_1 and $t_0$and are calculated in MENDEL, using a pedigree peeling algorithm[7]. When predicting future risks, this involves performing this calculation repeatedly at several time-points. However, under the standard assumption in pedigree likelihood calculations, the phenotype of the proband is conditionally independent of those of relatives given the genotypes of the relatives[28] . Thus:

$$\frac{\sum_G P(D(t_1)|G, D_R, \underline{z}, \underline{\theta}) P(G, D(t_0), D_R, \underline{z}, \underline{\theta})}{P(D(t_0), D_R, \underline{z}, \underline{\theta})}$$

which can be re-written as:

$$\sum_G P\big(D(t_1)\big|G, \underline{z}, \underline{\theta}\big) P\big(G\big|D_R, D(t_0), \underline{z}, \underline{\theta}\big) \qquad (2)$$

where *G* is the full set of genotypes (including the full measured and unmeasured polygenic or major gene components) and

$$P\big(G\big|D_R, D(t_0), \underline{z}, \underline{\theta}\big) = \frac{P(G, D(t_0) D_R, \underline{z}, \underline{\theta})}{\sum_{G'} P(G', D(t_0), D_R, \underline{z}, \underline{\theta})}$$

Therefore, the risk prediction (expression (2)) can be performed by first calculating the genotype probabilities for the proband given the phenotypes at time t_0 (i.e. a single, time-consuming pedigree likelihood calculation) and then calculating the penetrance function for the proband at multiple time-points $P(D(t_1)|G, \underline{z}, \underline{\theta})$, which does not involve any pedigree likelihood calculations.

The risk calculations under the revised and original formulations are identical, but when calculating the remaining lifetime cancer risks used in the CanRisk tool (www.canrisk.org), there is a 50-90% reduction in computation time under this revised formulation, depending on the proband's age (Figure s4).

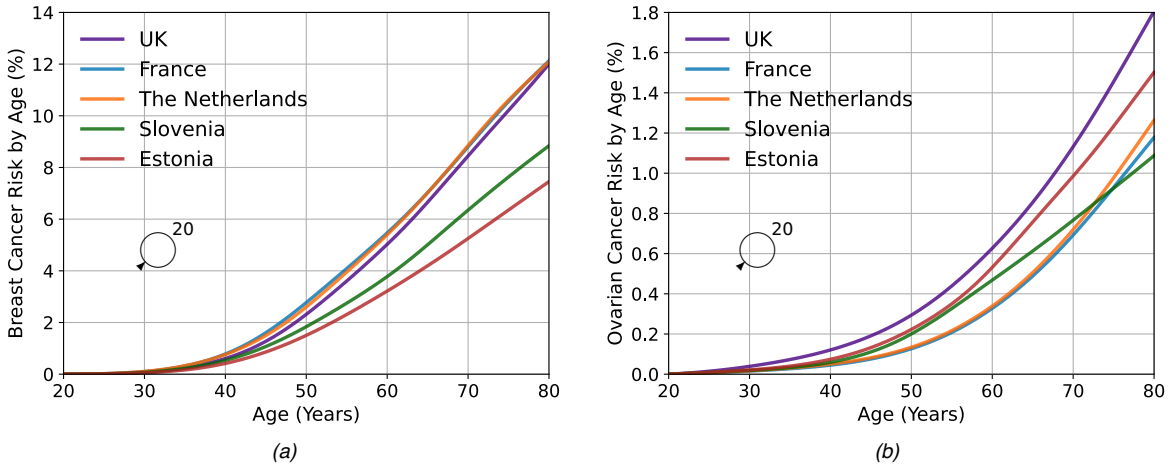


Figure s1. Predicted lifetime (age 20 to 80 years) breast and ovarian cancer risk by age for a female born in 1985 with unknown family history (ie average female in the population) comparing risks using incidences for the UK, France, the Netherlands, Slovenia, and Estonia. Figure (a) shows breast cancer risks, where risks for the UK, France, the Netherlands, Slovenia, and Estonia are 12.0%, 12.1%, 12.1%, 8.8% and 7.4%, respectively. Figure (b) shows ovarian cancer risks, where risks for the UK, France, the Netherlands, Slovenia, and Estonia are 1.8%, 1.2%, 1.3%, 1.1% and 1.5%, respectively.

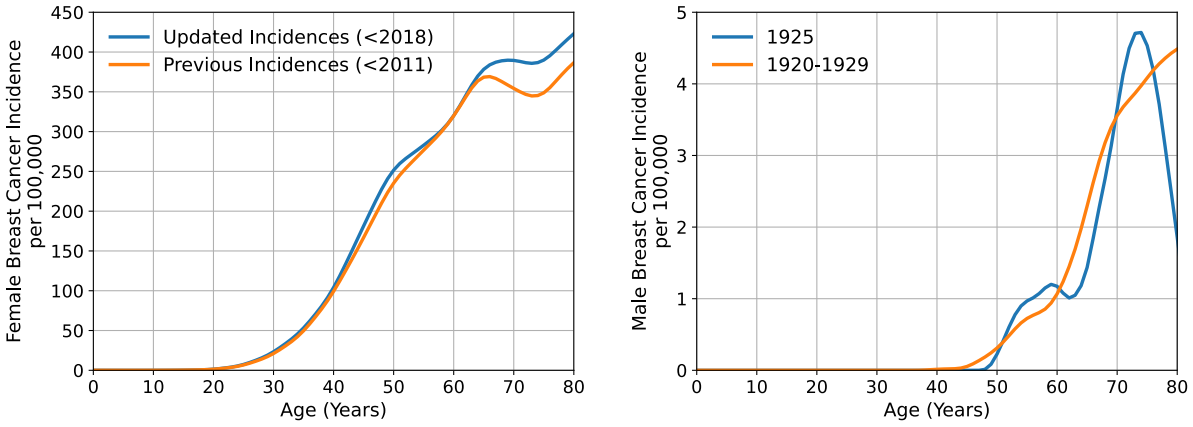


Figure s2. Smoothed Cohort-specific population incidences. Figure (a) shows female breast cancer incidences for the UK for those born in the 1980s for the previous incidences (using incidences up to and including 2010) and for the updated incidences (using incidences up to and including 2017), where both datasets use the average over the birth years in the cohort. Figure (b) shows male breast cancer incidences for Estonia for those born in the 1920s, using incidence from a single birth year to represent the cohort (labelled 1925) and using the average over the birth years in the cohort (labelled 1920-1929).

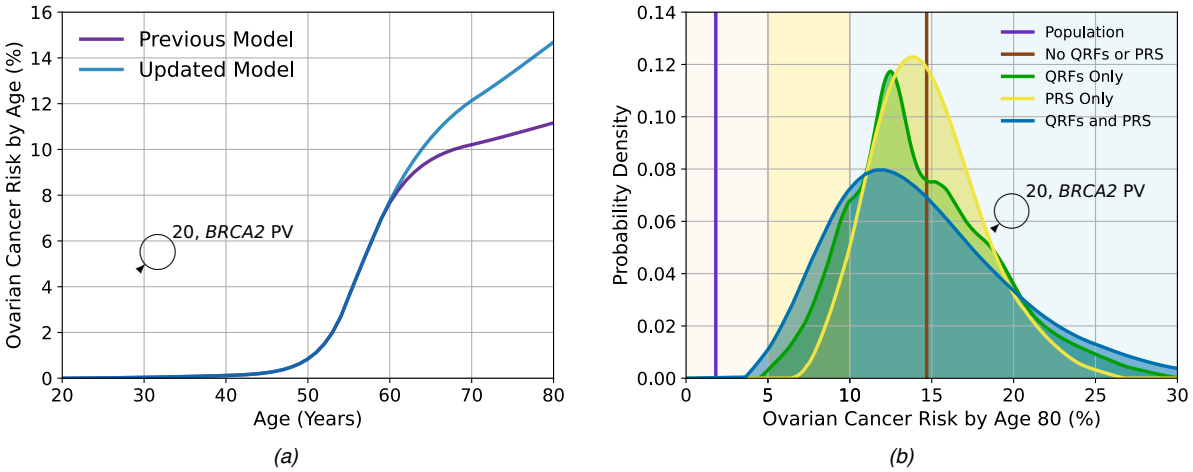


Figure s3. Revised epithelial ovarian cancer risks for a *BRCA2* pathogenic variant carrier with unknown family history using the updated *BRCA2* relative risks. Figure (a) shows the cumulative risk by age, while figure (b) shows the distribution of absolute by age 80 on the basis of the different predictors of risk (pathogenic variant status (PV), questionnaire-based risk factors (QRFs), mammographic density (MD), and PRS).

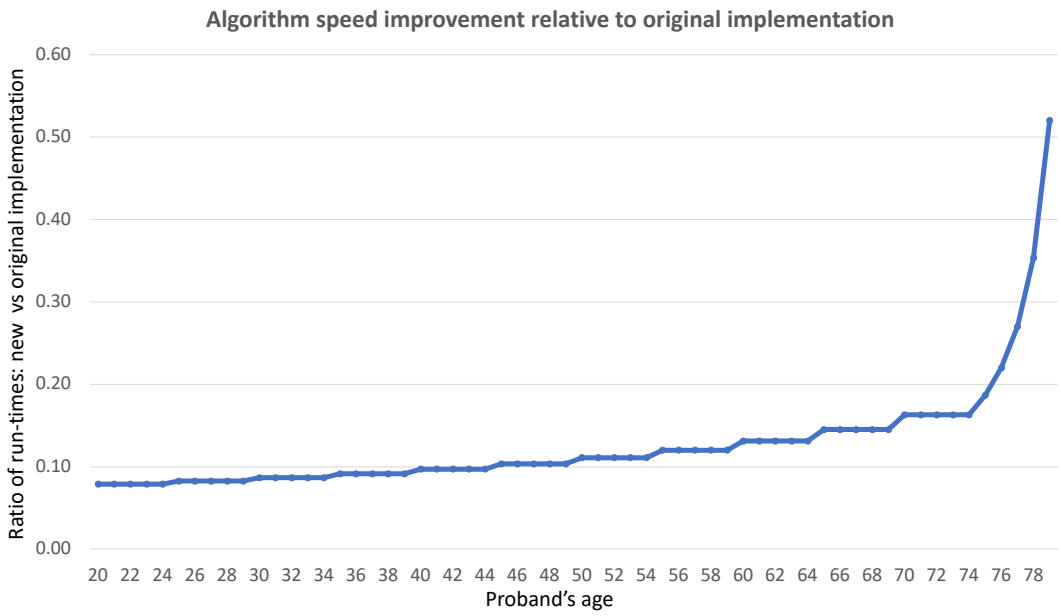


Figure s4. Ratio of run-times for calculating the remaining life-time risks in the CanRisk tool under the optimised algorithm compared to the original implementation.

New Susceptibility Genes

Breast Cancer: *BARD1*, *RAD51C* and *RAD51D*

Table s1. Predicted 10-year (age 40 to 50 years) and lifetime (age 20 to 80 years) breast cancer risk for a female born in 1985 with unknown family history and for a female with a mother affected at age 50. The columns labelled "Risk" contain risks in the absence of information about questionnaire-based risk factors (QRF), mammographic density (MD) or a polygenic risk score (PRS). The other columns show the distribution of females based on these risk factors in the risk categories defined in the NICE familial breast cancer guidelines [29]: 1) near-population risk, shaded pink ($< 17\%$ lifetime risk; $< 3\%$ 10-year risk), 2) moderate risk, shaded yellow ($\geq 17\%$ and $< 30\%$ lifetime risk; $\geq 3\%$ and $< 8\%$ 10-year risk) and 3) high risk, shaded blue ($\geq 30\%$ lifetime risk; $\geq 8\%$ 10-year risk). Column headings are shaded the same colours as the corresponding lines in Figure 2. Predictions are based on UK cancer incidences, assuming the population distributions of QRFs and MD.

Family History	Risk Horizon	PV Screening		QRF			QRF & MD			PRS			QRF & PRS			QRF, MD & PRS		
			Risk	Pop	Mod	High	Pop	Mod	High	Pop	Mod	High	Pop	Mod	High	Pop	Mod	High
Unknown	10-Year	Untested	1.7	96.4	3.6	0	94.9	5.1	0	90	10	0	90.6	9.2	0.2	89.2	10.6	0.3
		BARD1	3.6	48.8	50.5	0.6	48.8	49.8	1.4	46.5	50	3.5	56.5	39.5	4	57.8	37	5.2
		RAD51C	3.4	57.3	42.3	0.4	54	44.9	1.1	50.8	46.5	2.7	60.2	36.6	3.2	61	34.7	4.4
		RAD51D	3.1	67.3	32.5	0.2	61.1	38.2	0.7	56.5	41.6	1.8	64.9	32.8	2.3	65.2	31.4	3.4
		No PV	1.6	97.1	2.9	0	95.8	4.2	0	91.5	8.4	0	91.9	8	0.1	90.5	9.3	0.2
	Lifetime	Untested	12	92.1	7.7	0.2	88.8	10.7	0.5	87.5	12.3	0.2	85.2	13.7	1.1	83.4	14.9	1.7
		BARD1	23.5	12.5	77.1	10.4	21.8	62.8	15.3	19.8	62.1	18.1	30.8	49.4	19.8	33.9	44.1	21.9
		RAD51C	22.3	18.9	72.7	8.4	27.5	60.3	12.2	24.8	61	14.2	35.6	47.8	16.6	38.3	42.9	18.8
		RAD51D	20.8	29.7	63.9	6.4	35.8	55.3	8.9	32.4	57.7	9.9	42.4	44.7	12.9	44.3	40.7	15
		No PV	11.6	92.9	7	0.1	90.1	9.4	0.5	89.1	10.7	0.1	86.7	12.4	1	84.8	13.7	1.5
Mother affected at age 50	10-Year	Untested	3.5	51.8	47.6	0.5	50.6	48.1	1.3	53.7	45.7	0.6	63.4	35.1	1.5	63.5	34.1	2.4
		BARD1	6.4	1.1	87.2	11.7	10	71.9	18.1	12.5	69	18.5	23.3	60.3	16.4	28.6	53.1	18.3
		RAD51C	6.1	2	88.5	9.5	12.2	73	14.8	15	69.7	15.4	26.4	59.5	14.1	31.4	52.5	16.1
		RAD51D	5.6	3.8	88.9	7.3	15.8	73	11.1	18.7	69.5	11.8	30.8	57.8	11.4	35.4	51.2	13.4
		No PV	3.1	68.7	31.1	0.3	62.2	36.9	0.8	64.8	34.9	0.3	72	26.9	1.1	71.3	26.9	1.8
	Lifetime	Untested	19.3	40.6	54.6	4.8	43.5	50.1	6.3	47.1	50.3	2.5	54.6	39.2	6.2	55.0	36.9	8.1
		BARD1	33.7	0.1	37.9	62	3.4	37.4	59.3	2.3	43.2	54.6	7.8	44.2	47.9	12	40.2	47.8
		RAD51C	32.2	0.2	47.4	52.4	4.1	43.3	52.6	3.4	48.8	47.8	10	47.3	42.7	14.5	42.3	43.2
		RAD51D	30.4	0.5	59.3	40.2	5.5	50.7	43.8	5.6	55.5	38.9	13.6	50.3	36	18.3	44.4	37.3
		No PV	17.9	55	41.3	3.8	53.5	41.6	4.9	57.8	40.5	1.7	62.6	32.5	4.9	62.1	31.3	6.6

Ovarian cancer: *PALB2*

Table s2. Predicted ovarian cancer risk to age 50 (age 20 to 50 years) and lifetime risk (age 20 to 80 years) for a female born in 1985 with unknown family history and for a female with a mother affected at age 50. The columns labelled "Risk" contain risks in the absence of information about risk factors (RF) or a polygenic risk score (PRS). The other columns show the distribution of females based on these risk factors falling into risk categories defined as: 1) near-population risk, shaded pink ($< 5\%$ lifetime risk; $< 3\%$ risk to age 50), 2) moderate risk, shaded yellow ($\geq 5\%$ and $< 10\%$ lifetime risk; $\geq 3\%$ and $< 5\%$ risk to age 50) and 3) high risk, shaded blue ($\geq 10\%$ lifetime risk; $\geq 5\%$ risk to age 50). Column headings are shaded the same colours as the corresponding lines in Figure 1. Predictions are based on UK cancer incidences, assuming the population distributions of QRFs and MD.

Family History	Risk Horizon	PV Screening		RF			PRS			RF & PRS		
			Risk	Pop	Mod	High	Pop	Mod	High	Pop	Mod	High
Unknown	Risk to age 50	Untested	0.3	99.9	0	0.1	100	0	0	100	0	0
		PALB2	0.8	99.9	0	0.1	100	0	0	100	0	0
		No PV	0.3	99.9	0	0.1	100	0	0	100	0	0
	Lifetime	Untested	1.8	99.9	0.1	0.1	100	0	0	99.6	0.4	0
		PALB2	5	61.8	37.3	1	55.8	44.1	0.1	62.4	34.8	2.7
		No PV	1.7	99.9	0	0.1	100	0	0	99.7	0.3	0
Mother affected at age 50	Risk to age 50	Untested	1	99.9	0	0.1	100	0	0	100	0	0
		PALB2	1.5	98.4	1.6	0.1	100	0	0	97.5	2.4	0.1
		No PV	0.5	99.9	0	0.1	100	0	0	100	0	0
	Lifetime	Untested	5.1	57.6	41.8	0.6	52.9	47.0	0	60.5	38.3	1.2
		PALB2	9.6	6.1	58.7	35.2	0.4	66.0	34.6	11.4	56.0	32.6
		No PV	3.3	91.7	8.2	0.1	98.6	1.3	0	90.5	9.4	0.1

Tumour Pathology Subtypes

AGE	GENERAL POPULATION	*BRCA1*	*BRCA2*	*PALB2*	*CHEK2*	*ATM*	*BARD1*	*RAD51C*	*RAD51D*
20	0.4615	0.8201	0.1881	0.5316	0.2884	0.2366	0.7706	0.7471	0.6255
21	0.4615	0.8201	0.1881	0.5316	0.2884	0.2366	0.7706	0.7471	0.6255
22	0.4615	0.8201	0.1881	0.5316	0.2884	0.2366	0.7706	0.7471	0.6255
23	0.4615	0.8201	0.1881	0.5316	0.2884	0.2366	0.7706	0.7471	0.6255
24	0.4614	0.8201	0.1881	0.5316	0.2884	0.2366	0.7706	0.7471	0.6255
25	0.4606	0.8201	0.1881	0.5316	0.2884	0.2366	0.7706	0.7471	0.6255
26	0.459	0.8201	0.1881	0.5218	0.2813	0.2296	0.7633	0.7394	0.6159
27	0.4567	0.8201	0.1881	0.512	0.2743	0.2227	0.7558	0.7314	0.6061
28	0.4539	0.8201	0.1881	0.5021	0.2674	0.216	0.748	0.7232	0.596
29	0.4504	0.8201	0.1881	0.4921	0.2605	0.2093	0.74	0.7147	0.5858
30	0.4444	0.8201	0.1881	0.482	0.2537	0.2027	0.7318	0.706	0.5754
31	0.435	0.8201	0.1881	0.4718	0.2469	0.1961	0.7232	0.697	0.5647
32	0.4228	0.8201	0.1881	0.4615	0.2402	0.1897	0.7144	0.6877	0.5539
33	0.4093	0.8201	0.1881	0.4513	0.2337	0.1835	0.7055	0.6784	0.5431
34	0.3953	0.8196	0.1883	0.441	0.2272	0.1774	0.6963	0.6688	0.5322
35	0.3804	0.8171	0.1891	0.4307	0.2207	0.1713	0.6868	0.6588	0.521
36	0.3638	0.8117	0.1908	0.4202	0.2142	0.1653	0.677	0.6486	0.5096
37	0.3461	0.804	0.1932	0.4096	0.2079	0.1594	0.6668	0.638	0.4979
38	0.3289	0.7947	0.1962	0.3989	0.2015	0.1536	0.6562	0.627	0.4859
39	0.3139	0.7849	0.1993	0.3881	0.1951	0.1478	0.6451	0.6156	0.4737
40	0.3013	0.775	0.2024	0.3771	0.1888	0.142	0.6336	0.6038	0.4611
41	0.2899	0.7652	0.2055	0.3661	0.1826	0.1364	0.6217	0.5915	0.4483
42	0.2785	0.7559	0.2085	0.3549	0.1763	0.1308	0.6093	0.5788	0.4353
43	0.2676	0.7482	0.2109	0.3469	0.1717	0.1269	0.6008	0.5703	0.4266
44	0.2583	0.7426	0.213	0.3389	0.1671	0.123	0.5922	0.5615	0.4179
45	0.2511	0.7388	0.2157	0.3309	0.1625	0.1191	0.5834	0.5525	0.409
46	0.2451	0.7357	0.2197	0.3227	0.1579	0.1153	0.5743	0.5433	0.4001
47	0.2395	0.7318	0.2253	0.3146	0.1533	0.1115	0.565	0.5338	0.391
48	0.2343	0.7273	0.232	0.3094	0.1503	0.1092	0.5597	0.5285	0.386
49	0.2301	0.7224	0.2392	0.3043	0.1473	0.1069	0.5544	0.523	0.3809
50	0.2269	0.7176	0.2463	0.2991	0.1443	0.1045	0.549	0.5175	0.3759
51	0.2243	0.7127	0.2535	0.294	0.1413	0.1022	0.5435	0.512	0.3708
52	0.2217	0.7082	0.2602	0.2888	0.1383	0.0999	0.538	0.5064	0.3657
53	0.2192	0.7043	0.2658	0.2852	0.1363	0.0984	0.5344	0.5027	0.3624
54	0.2167	0.7008	0.2697	0.2817	0.1343	0.0968	0.5307	0.499	0.3592
55	0.2139	0.695	0.2717	0.2781	0.1323	0.0953	0.5271	0.4953	0.3559
56	0.2103	0.6852	0.2725	0.2745	0.1303	0.0937	0.5234	0.4916	0.3526
57	0.2058	0.6713	0.273	0.2709	0.1283	0.0922	0.5197	0.4878	0.3493
58	0.2009	0.6547	0.2736	0.2678	0.1266	0.0908	0.5164	0.4845	0.3464
59	0.1961	0.6371	0.2743	0.2646	0.1249	0.0895	0.5131	0.4812	0.3435
60	0.1911	0.6193	0.2749	0.2614	0.1232	0.0882	0.5098	0.4778	0.3407
61	0.1858	0.6017	0.2756	0.2583	0.1214	0.0868	0.5064	0.4745	0.3378
62	0.1803	0.5851	0.2762	0.2551	0.1197	0.0855	0.5031	0.4711	0.3349
63	0.1751	0.5712	0.2767	0.2524	0.1185	0.0844	0.5	0.468	0.3322
64	0.1707	0.5617	0.2771	0.2498	0.1172	0.0833	0.4969	0.4649	0.3295
65	0.1677	0.5571	0.2773	0.2471	0.1159	0.0823	0.4938	0.4618	0.3268
66	0.1656	0.5562	0.2773	0.2445	0.1146	0.0812	0.4907	0.4587	0.3242
67	0.1641	0.5562	0.2773	0.2418	0.1133	0.0801	0.4876	0.4556	0.3215
68	0.1631	0.5562	0.2773	0.2392	0.1121	0.079	0.4845	0.4524	0.3188
69	0.1625	0.5562	0.2773	0.2366	0.1108	0.078	0.4813	0.4493	0.3161
70	0.162	0.5562	0.2773	0.234	0.1096	0.077	0.4782	0.4461	0.3135
71	0.1613	0.5562	0.2773	0.2314	0.1083	0.0759	0.475	0.443	0.3108
72	0.1601	0.5562	0.2773	0.2287	0.1071	0.0749	0.4718	0.4398	0.3081
73	0.1584	0.5562	0.2773	0.2263	0.106	0.0739	0.4688	0.4368	0.3056
74	0.1564	0.5562	0.2773	0.2239	0.1049	0.073	0.4658	0.4338	0.3031
75	0.1538	0.5562	0.2773	0.2215	0.1038	0.072	0.4628	0.4308	0.3006
76	0.1507	0.5562	0.2773	0.219	0.1026	0.0711	0.4597	0.4278	0.2981
77	0.1474	0.5562	0.2773	0.2166	0.1015	0.0701	0.4566	0.4248	0.2955
78	0.1446	0.5562	0.2773	0.2142	0.1004	0.0692	0.4536	0.4217	0.293
79	0.1434	0.5562	0.2773	0.2117	0.0993	0.0683	0.4505	0.4186	0.2905

Table s3: Age-specific proportion of oestrogen receptor-negative tumours among all female breast cancer tumours in the general population and carriers of pathogenic variants in the breast cancer susceptibility genes used in the BOADICEA model.

AGE	GENERAL POPULATION	*BRCA1*	*BRCA2*	*PALB2*	*CHEK2*	*ATM*	*BARD1*	*RAD51C*	*RAD51D*
20	0.6582	0.8799	0.7586	0.6066	0.3768	0.5979	0.8195	0.8094	0.8489
21	0.6582	0.8799	0.7586	0.6066	0.3768	0.5979	0.8195	0.8094	0.8489
22	0.6582	0.8799	0.7586	0.6066	0.3768	0.5979	0.8195	0.8094	0.8489
23	0.6582	0.8799	0.7586	0.6066	0.3768	0.5979	0.8195	0.8094	0.8489
24	0.6579	0.8799	0.7586	0.6066	0.3768	0.5979	0.8195	0.8094	0.8489
25	0.6566	0.8799	0.7586	0.6066	0.3768	0.5979	0.8195	0.8094	0.8489
26	0.6539	0.8799	0.7586	0.6047	0.3766	0.596	0.8183	0.8082	0.8479
27	0.6499	0.8799	0.7586	0.6027	0.3763	0.594	0.8171	0.8069	0.8468
28	0.6451	0.8799	0.7586	0.6006	0.3759	0.5919	0.8158	0.8055	0.8457
29	0.6401	0.8799	0.7586	0.5984	0.3754	0.5897	0.8144	0.8041	0.8445
30	0.6349	0.8799	0.7586	0.5961	0.3748	0.5874	0.813	0.8026	0.8433
31	0.6299	0.8799	0.7586	0.5937	0.3741	0.5849	0.8114	0.801	0.8419
32	0.6251	0.8799	0.7586	0.5911	0.3732	0.5824	0.8098	0.7993	0.8405
33	0.6211	0.8799	0.7586	0.5888	0.3726	0.58	0.8083	0.7977	0.8392
34	0.6181	0.8799	0.7586	0.5863	0.3719	0.5775	0.8067	0.7961	0.8378
35	0.6156	0.8799	0.7586	0.5837	0.371	0.5749	0.805	0.7943	0.8364
36	0.6128	0.8799	0.7586	0.5809	0.3699	0.5721	0.8032	0.7925	0.8348
37	0.6091	0.8799	0.7586	0.578	0.3688	0.5691	0.8013	0.7905	0.8331
38	0.6046	0.8799	0.7586	0.5747	0.3673	0.5658	0.7992	0.7882	0.8312
39	0.5999	0.8799	0.7586	0.5712	0.3655	0.5623	0.7968	0.7858	0.8292
40	0.5952	0.8799	0.7586	0.5674	0.3636	0.5585	0.7943	0.7832	0.827
41	0.5905	0.8799	0.7586	0.5634	0.3614	0.5545	0.7917	0.7804	0.8247
42	0.586	0.8799	0.7586	0.5591	0.359	0.5501	0.7888	0.7774	0.8221
43	0.5823	0.8799	0.7586	0.5588	0.3604	0.5499	0.7886	0.7772	0.822
44	0.5798	0.8799	0.7586	0.5585	0.3617	0.5496	0.7884	0.777	0.8218
45	0.5785	0.8799	0.7586	0.5583	0.3631	0.5493	0.7882	0.7768	0.8216
46	0.5783	0.8799	0.7586	0.558	0.3644	0.549	0.788	0.7766	0.8215
47	0.5783	0.8799	0.7586	0.5576	0.3657	0.5487	0.7878	0.7764	0.8213
48	0.5783	0.8799	0.7586	0.5602	0.3698	0.5513	0.7895	0.7782	0.8228
49	0.5783	0.8799	0.7586	0.563	0.374	0.554	0.7914	0.7801	0.8244
50	0.5783	0.8799	0.7586	0.5658	0.3784	0.5569	0.7933	0.7821	0.8261
51	0.5783	0.8799	0.7586	0.5687	0.3828	0.5598	0.7952	0.7841	0.8278
52	0.5783	0.8799	0.7586	0.5718	0.3875	0.5629	0.7972	0.7862	0.8295
53	0.5783	0.8799	0.7586	0.5749	0.3922	0.566	0.7993	0.7884	0.8313
54	0.5783	0.8799	0.7586	0.5781	0.397	0.5692	0.8014	0.7905	0.8332
55	0.5783	0.8799	0.7586	0.5814	0.4019	0.5726	0.8035	0.7928	0.8351
56	0.5783	0.8799	0.7586	0.5848	0.407	0.576	0.8057	0.7951	0.837
57	0.5783	0.8799	0.7586	0.5883	0.4121	0.5795	0.808	0.7974	0.8389
58	0.5783	0.8799	0.7586	0.5915	0.4171	0.5827	0.81	0.7996	0.8407
59	0.5783	0.8799	0.7586	0.5948	0.4221	0.5861	0.8121	0.8018	0.8426
60	0.5783	0.8799	0.7586	0.5982	0.4273	0.5895	0.8143	0.804	0.8444
61	0.5783	0.8799	0.7586	0.6017	0.4326	0.593	0.8165	0.8063	0.8463
62	0.5783	0.8799	0.7586	0.6053	0.438	0.5966	0.8187	0.8086	0.8483
63	0.5783	0.8799	0.7586	0.6073	0.4417	0.5986	0.8199	0.8099	0.8493
64	0.5783	0.8799	0.7586	0.6093	0.4455	0.6006	0.8212	0.8111	0.8504
65	0.5783	0.8799	0.7586	0.6113	0.4494	0.6027	0.8224	0.8125	0.8515
66	0.5783	0.8799	0.7586	0.6134	0.4533	0.6048	0.8237	0.8138	0.8526
67	0.5783	0.8799	0.7586	0.6155	0.4572	0.6069	0.825	0.8152	0.8537
68	0.5783	0.8799	0.7586	0.6176	0.4611	0.609	0.8262	0.8164	0.8548
69	0.5783	0.8799	0.7586	0.6196	0.4651	0.6111	0.8275	0.8178	0.8559
70	0.5783	0.8799	0.7586	0.6218	0.4691	0.6132	0.8288	0.8191	0.857
71	0.5783	0.8799	0.7586	0.624	0.4731	0.6154	0.8301	0.8205	0.8581
72	0.5783	0.8799	0.7586	0.6262	0.4773	0.6177	0.8315	0.8219	0.8593
73	0.5783	0.8799	0.7586	0.6283	0.4812	0.6197	0.8327	0.8232	0.8603
74	0.5783	0.8799	0.7586	0.6303	0.4851	0.6218	0.8339	0.8245	0.8614
75	0.5783	0.8799	0.7586	0.6324	0.4892	0.624	0.8352	0.8258	0.8625
76	0.5783	0.8799	0.7586	0.6346	0.4933	0.6262	0.8365	0.8271	0.8636
77	0.5783	0.8799	0.7586	0.6368	0.4974	0.6284	0.8378	0.8285	0.8647
78	0.5783	0.8799	0.7586	0.6391	0.5016	0.6307	0.8391	0.8299	0.8659
79	0.5783	0.8799	0.7586	0.6415	0.5059	0.6331	0.8405	0.8313	0.8671

Table s4: Age-specific proportion of triple-negative tumours among female oestrogen receptor-negative breast cancer tumours in the general population and carriers of pathogenic variants in the breast cancer susceptibility genes used in the BOADICEA model.

References

1. Antoniou AC, Pharoah PP, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer* 2004;91(8):1580-90. doi: 10.1038/sj.bjc.6602175 [doi];6602175 [pii]
2. Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA, Risch HA, Eyfjord JE, Hopper JL, Southey MC, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tryggvadottir L, Syrjakoski K, Kallioniemi OP, Eerola H, Nevanlinna H, Pharoah PD, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* 2008;98(8):1457-66. doi: 6604305 [pii];10.1038/sj.bjc.6604305 [doi]
3. Jervis S, Song H, Lee A, Dicks E, Harrington P, Baynes C, Manchanda R, Easton DF, Jacobs I, Pharoah PP, Antoniou AC. A risk prediction algorithm for ovarian cancer incorporating BRCA1, BRCA2, common alleles and other familial effects. *J Med Genet* 2015;52(7):465-75. doi: 10.1136/jmedgenet-2015-103077
4. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, Babb de Villiers C, Izquierdo A, Simard J, Schmidt MK, Walter FM, Chatterjee N, Garcia-Closas M, Tischkowitz M, Pharoah P, Easton DF, Antoniou AC. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med* 2019;21(8):1708-18. doi: 10.1038/s41436-018-0406-9 [published Online First: 2019/01/16]
5. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol* 2001;21(1):1-18. doi: 10.1002/gepi.1014 [pii];10.1002/gepi.1014 [doi]
6. Lee A, Yang X, Tyrer J, Gentry-Maharaj A, Ryan A, Mavaddat N, Cunningham AP, Carver T, Archer S, Leslie G, Kalsi J, Gaba F, Manchanda R, Gayther SA, Ramus SJ, Walter FM, Tischkowitz M, Jacobs I, Menon U, Easton DF, Pharoah PPD, Antoniou AC. A Comprehensive Epithelial Tubo-Ovarian Cancer Risk Prediction Model Incorporating Genetic and Epidemiological Risk Factors. *J Med Genet* 2021:In Press.
7. Lange K, Weeks D, Boehnke M. Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol* 1988;5(6):471-72. doi: 10.1002/gepi.1370050611 [doi]
8. Cannings C, Thompson EA, Skolnick MH. Probability functions on complex pedigrees. *Advances in Applied Probability* 1978;10(1):26-61. doi: 10.2307/1426718 [published Online First: 2016/07/01]
9. Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, Pooley KA, Parsons MT, Fortuno C, Wang Q, Bolla MK, Dennis J, Keeman R, Alonso MR, Álvarez N, Herraez B, Fernandez V, Núñez-Torres R, Osorio A, Valcich J, Li M, Törngren T, Harrington PA, Baynes C, Conroy DM, Decker B, Fachal L, Mavaddat N, Ahearn T, Aittomäki K, Antonenkova NN, Arnold N, Arveux P, Ausems MGEM, Auvinen P, Becher H, Beckmann MW, Behrens S, Bermisheva M, Białkowska K, Blomqvist C, Bogdanova NV, Bogdanova-Markov N, Bojesen SE, Bonanni B, Børresen-Dale AL, Brauch H, Bremer M, Briceno I, Brüning T, Burwinkel B, Cameron DA, Camp NJ, Campbell A, Carracedo A, Castelao JE, Cessna MH, Chanock SJ, Christiansen H, Collée JM, Cordina-Duverger E, Cornelissen S, Czene K, Dörk T, Ekici AB, Engel C, Eriksson M, Fasching PA, Figueroa J, Flyger H, Försti A, Gabrielson M, Gago-Dominguez M, Georgoulias V, Gil F, Giles GG, Glendon G, Garcia EBG, Alnæs GIG, Guénel P,

Hadjisavvas A, Haeberle L, Hahnen E, Hall P, Hamann U, Harkness EF, Hartikainen JM, Hartman M, He W, Heemskerk-Gerritsen BAM, Hillemanns P, Hogervorst FBL, Hollestelle A, Ho WK, Hooning MJ, Howell A, Humphreys K, Idris F, Jakubowska A, Jung A, Kapoor PM, Kerin MJ, Khusnutdinova E, Kim SW, Ko YD, Kosma VM, Kristensen VN, Kyriacou K, Lakeman IMM, Lee JW, Lee MH, Li J, Lindblom A, Lo WY, Loizidou MA, Lophatananon A, Lubiński J, MacInnis RJ, Madsen MJ, Mannermaa A, Manoochehri M, Manoukian S, Margolin S, Martinez ME, Maurer T, Mavroudis D, McLean C, Meindl A, Mensenkamp AR, Michailidou K, Miller N, Mohd Taib NA, Muir K, Mulligan AM, Nevanlinna H, Newman WG, Nordestgaard BG, Ng PS, Oosterwijk JC, Park SK, Park-Simon TW, Perez JIA, Peterlongo P, Porteous DJ, Prajzendanc K, Prokofyeva D, Radice P, Rashid MU, Rhenius V, Rookus MA, Rüdiger T, Saloustros E, Sawyer EJ, Schmutzler RK, Schneeweiss A, Schürmann P, Shah M, Sohn C, Southey MC, Surowy H, Suvanto M, Thanasitthichai S, Tomlinson I, Torres D, Truong T, Tzardi M, Valova Y, van Asperen CJ, Van Dam RM, van den Ouweland AMW, van der Kolk LE, van Veen EM, Wendt C, Williams JA, Yang XR, Yoon SY, Zamora MP, Evans DG, de la Hoya M, Simard J, Antoniou AC, Borg Å, Andrulis IL, Chang-Claude J, García-Closas M, Chenevix-Trench G, Milne RL, Pharoah PDP, Schmidt MK, Spurdle AB, Vreeswijk MPG, Benitez J, Dunning AM, Kvist A, Teo SH, Devilee P, Easton DF, Consortium BCA. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* 2021;384(5):428-39. doi: 10.1056/NEJMoa1913948 [published Online First: 2021/01/20]

10. Dorling L, Carvalho S, Allen J, Parsons MT, Fortuno C, González-Neira A, Heijl SM, Adank MA, Ahearn TU, Andrulis IL, Auvinen P, Becher H, Beckmann MW, Behrens S, Bermisheva M, Bogdanova NV, Bojesen SE, Bolla MK, Bremer M, Briceno I, Camp NJ, Campbell A, Castelao JE, Chang-Claude J, Chanock SJ, Chenevix-Trench G, Collaborators N, Collée JM, Czene K, Dennis J, Dörk T, Eriksson M, Evans DG, Fasching PA, Figueroa J, Flyger H, Gabrielson M, Gago-Dominguez M, García-Closas M, Giles GG, Glendon G, Guénel P, Gündert M, Hadjisavvas A, Hahnen E, Hall P, Hamann U, Harkness EF, Hartman M, Hogervorst FBL, Hollestelle A, Hoppe R, Howell A, Investigators k, Investigators S, Jakubowska A, Jung A, Khusnutdinova E, Kim S-W, Ko Y-D, Kristensen VN, Lakeman IMM, Li J, Lindblom A, Loizidou MA, Lophatananon A, Lubiński J, Luccarini C, Madsen MJ, Mannermaa A, Manoochehri M, Margolin S, Mavroudis D, Milne RL, Mohd Taib NA, Muir K, Nevanlinna H, Newman WG, Oosterwijk JC, Park SK, Peterlongo P, Radice P, Saloustros E, Sawyer EJ, Schmutzler RK, Shah M, Sim X, Southey MC, Surowy H, Suvanto M, Tomlinson I, Torres D, Truong T, van Asperen CJ, Waltes R, Wang Q, Yang XR, Pharoah PDP, Schmidt MK, Benitez J, Vroling B, Dunning AM, Teo SH, Kvist A, de la Hoya M, Devilee P, Spurdle AB, Vreeswijk MPG, Easton DF. Breast cancer risks associated with missense variants in breast cancer susceptibility genes. *medRxiv* 2021:2021.09.02.21262369. doi: 10.1101/2021.09.02.21262369

11. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, Jervis S, van Leeuwen FE, Milne RL, Andrieu N, Goldgar DE, Terry MB, Rookus MA, Easton DF, Antoniou AC, Brca, Consortium BC, McGuffog L, Evans DG, Barrowdale D, Frost D, Adlard J, Ong KR, Izatt L, Tischkowitz M, Eeles R, Davidson R, Hodgson S, Ellis S, Nogues C, Lasset C, Stoppa-Lyonnet D, Fricker JP, Faivre L, Berthet P, Hooning MJ, van der Kolk LE, Kets CM, Adank MA, John EM, Chung WK, Andrulis IL, Southey M, Daly MB, Buys SS, Osorio A, Engel C, Kast K, Schmutzler RK, Caldes T, Jakubowska A,

Simard J, Friedlander ML, McLachlan SA, Machackova E, Foretova L, Tan YY, Singer CF, Olah E, Gerdes AM, Arver B, Olsson H. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* 2017;317(23):2402-16. doi: 10.1001/jama.2017.7112
12. Lee AJ, Cunningham AP, Kuchenbaecker KB, Mavaddat N, Easton DF, Antoniou AC, Consortium of Investigators of Modifiers of B, Breast Cancer Association C. BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *Br J Cancer* 2014;110(2):535-45. doi: 10.1038/bjc.2013.730
13. Statistics Netherlands. [Available from: www.cbs.nl.
14. Netherlands Cancer Registry. [Available from: www.iknl.nl.
15. Ferlay J, Parkin DM, Curado MP, Bray F, Edwards B, Shin HR, Forman D. Cancer Incidence in Five Continents, Volumes I to IX: IARC CancerBase No. 9 [Internet]: International Agency for Research on Cancer, 2010.
16. Ferlay J, Bray F, Steliarova-Foucher E, Forman D. Cancer Incidence in Five Continents, CI5plus. IARC CancerBase No. 9: Lyon: International Agency for Research on Cancer, 2014.
17. Defossez G, Le Guyader-Peyrou S, Uhry Z, Grosclaude P, Colonna M, Dantony E, Delafosse P, Molinié F, Woronoff A-S, Bouvier A-M, Bossard N, Remontet L, Monnereau A. Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018 - Volume 1: Tumeurs solides : Étude à partir des registres des cancers du réseau Francim. *Santé publique France* 2019:372.
18. Slovenian Cancer Registry. [Available from: https://www.onko-i.si.
19. The Estonian National Institute for Health Development. Age-specific incidence rate of malignant neoplasms per 100 000 inhabitants by site and sex. [updated [Last Update: 2020 September 12]. Available from: https://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__02Haigestumus__04PahaloomulisedKasvajad/PK30.px/.
20. Office for National Statistics UK. Cancer Statistics Registrations, England, Series MB1 2017 [updated [Last Update: 2019-05-16]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland accessed August 23 2021.
21. Danckert B, Ferlay J, Engholm G, Hansen H, Johannesen T, Khan S, Køtlum J, Ólafsdóttir E, Schmidt L, Virtanen A, Storm H. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 8.2 (26.03.2019). Association of the Nordic Cancer Registries. Danish Cancer Society.; 2019 [Available from: http://www.ancr.nu accessed August 18 2021.
22. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, Kotlum JE, Olafsdottir E, Pukkala E, Storm HH. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. *Acta Oncol* 2010;49(5):725-36. doi: 10.3109/02841861003782017 [doi]
23. Australian Institute of Health and Welfare (AIHW). Australian Cancer Incidence and Mortality (ACIM) Books, CAN 122: AIHW: Canberra; 2021 [updated 08 Jun 2021. accessed August 22 2021.
24. Surveillance Epidemiology and End Results (SEER) Program. SEER*Stat Database: Incidence - SEER 21 Regs Research Data, Nov 2020 - Linked To County Attributes -

Total U.S. 2020 [updated 2020. Available from: www.seer.cancer.gov accessed August 18 2021.
25. Lewis C. Personal Communication: New Zealand Cancer Incidences, 2021.
26. New Zealand Ministry of Health. Cancer Statistis. In: Health Mo, ed., 2021.
27. Statistics Canada. Table 13-10-0111-01 Number and rates of new cases of primary cancer, by cancer type, age group and sex 2021 [updated [Last Update: 2020-10-23]. Available from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310011101 accessed August 22 2021.
28. Lange K. Mathematical and statistical methods for genetic analysis. New York: Springer 1997.
29. National Institute for Health and Care Excellence. Familial breast cancer: classification and care of people at risk of familial breast cancer and management of breast cancer and related risks in people with a family history of breast cancer (CG164) 2013 [updated 2019-11-20. Available from: http://www.nice.org.uk/CG164.