

## Article

# Combining Breast Cancer Risk Prediction Models

Zoe Guan <sup>1,\*</sup>, Theodore Huang <sup>2</sup>, Anne Marie McCarthy <sup>3</sup>, Kevin Hughes <sup>4</sup>, Alan Semine <sup>5</sup>, Hajime Uno <sup>6,7</sup>, Lorenzo Trippa <sup>6,8</sup>, Giovanni Parmigiani <sup>6,8</sup> and Danielle Braun <sup>6,8</sup>

<sup>1</sup> Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10017, USA

<sup>2</sup> Vertex Pharmaceuticals, Boston, MA 02210, USA

<sup>3</sup> Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>4</sup> Department of Surgery, Medical University of South Carolina, Charleston, SC 29425, USA

<sup>5</sup> Advanced Image Enhancement, Fall River, MA 02720, USA

<sup>6</sup> Department of Data Science, Dana-Farber Cancer Institute, Boston, MA 02115, USA

<sup>7</sup> Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

<sup>8</sup> Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

\* Correspondence: [guan@mskcc.org](mailto:guan@mskcc.org)

**Simple Summary:** BRCAPRO is a widely used breast cancer risk prediction model based on family history. A major limitation of this model is that it does not consider non-genetic risk factors. We expand BRCAPRO by combining it with another popular model, BCRAT, that uses mostly non-genetic risk factors, and show that the expanded model can achieve improvements in prediction accuracy over both BRCAPRO and BCRAT.

**Abstract:** Accurate risk stratification is key to reducing cancer morbidity through targeted screening and preventative interventions. Multiple breast cancer risk prediction models are used in clinical practice, and often provide a range of different predictions for the same patient. Integrating information from different models may improve the accuracy of predictions, which would be valuable for both clinicians and patients. BRCAPRO is a widely used model that predicts breast cancer risk based on detailed family history information. A major limitation of this model is that it does not consider non-genetic risk factors. To address this limitation, we expand BRCAPRO by combining it with another popular existing model, BCRAT (i.e., Gail), which uses a largely complementary set of risk factors, most of them non-genetic. We consider two approaches for combining BRCAPRO and BCRAT: (1) modifying the penetrance (age-specific probability of developing cancer given genotype) functions in BRCAPRO using relative hazard estimates from BCRAT, and (2) training an ensemble model that takes BRCAPRO and BCRAT predictions as input. Using both simulated data and data from Newton-Wellesley Hospital and the Cancer Genetics Network, we show that the combination models are able to achieve performance gains over both BRCAPRO and BCRAT. In the Cancer Genetics Network cohort, we show that the proposed BRCAPRO + BCRAT penetrance modification model performs comparably to IBIS, an existing model that combines detailed family history with non-genetic risk factors.

**Keywords:** BRCAPRO; BCRAT; model aggregation; ensemble learning; stacking



**Citation:** Guan, Z.; Huang, T.; McCarthy, A.M.; Hughes, K.; Semine, A.; Uno, H.; Trippa, L.; Parmigiani, G.; Braun, D. Combining Breast Cancer Risk Prediction Models. *Cancers* **2023**, *15*, 1090. <https://doi.org/10.3390/cancers15041090>

Academic Editors: Valentina Silvestri, Arianna Nicolussi and Laura Ottini

Received: 3 January 2023

Revised: 31 January 2023

Accepted: 2 February 2023

Published: 8 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

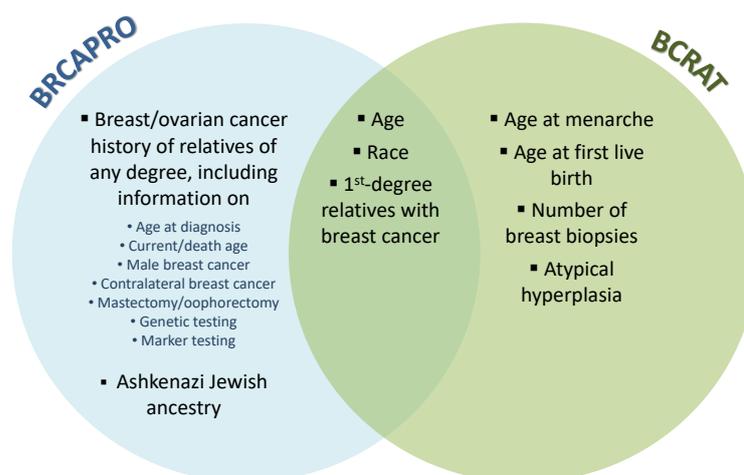
Breast cancer is the second most common cancer and the second leading cause of cancer death in women in the U.S. [1,2]. Identifying individuals at high risk is critical for guiding decisions about risk management and prevention, including screening, genetic counseling and testing, and preventative procedures. At least 24 breast cancer risk prediction models have been developed for clinical use [3]. These models estimate either an individual's risk of carrying a pathogenic mutation in a breast cancer susceptibility gene, an individual's future

risk of breast cancer, or both. They are based on a wide range of risk factors, methodologies, and study populations. Regression-based models, such as the Breast Cancer Risk Assessment Tool (BCRAT, also known as the Gail model) [4–7], use hormonal/reproductive risk factors (such as age at first live birth), history of benign disease, and simple summaries of family history. Other models, such as BRCAPRO [8], BOADICEA [9–11], and IBIS [12], are Mendelian models that use detailed family history information and principles of genetic inheritance. IBIS and BOADICEA [11] take into account hormonal/reproductive risk factors and history of benign disease. BRCAPRO currently does not use non-genetic risk factors, which limits its predictive accuracy [13,14]. Because it is widely adopted in clinical practice [15], adding non-genetic risk factors to the model is crucial. We address this problem by combining it with the widely used BCRAT model, which is based on a largely complementary set of risk factors, most of which are non-genetic.

BRCAPRO is a family history-based model that provides carrier probabilities for breast cancer susceptibility genes BRCA1 and BRCA2 as well as future risk estimates for invasive breast and ovarian cancer. It translates family history data into risk estimates using Mendelian laws of inheritance, Bayes' rule, and literature-based estimates of mutation prevalence and penetrance (age-specific probability of developing cancer given genotype). In contrast, BCRAT estimates an individual's future risk of invasive breast cancer based on a relative hazard model that includes age, hormonal and reproductive risk factors, history of benign disease, and first-degree family history of breast cancer. The relative hazard model was originally trained using case-control data from Caucasian women participating in a U.S. mammography screening program. It was later updated for African-American [5], Asian-American [6], and Hispanic [7] women. Although there is overlap in the inputs to BRCAPRO and BCRAT, the two models are largely complementary (Figure 1). BRCAPRO uses potentially extensive family history information, while BCRAT considers only first-degree female relatives with breast cancer. BCRAT considers several non-genetic risk factors that are not considered by BRCAPRO, including age at menarche, age at first live birth, and number of breast biopsies. A validation study in a large U.S. screening cohort found that 6-year risk predictions from BRCAPRO and BCRAT had a moderate correlation of 0.53 [16]. Because BRCAPRO and BCRAT embed different information, combining these models can potentially lead to accuracy gains.

There already exist models that incorporate both detailed family history information and non-genetic risk factors, namely, IBIS (also known as the Tyrer–Cuzick model) [12] and BOADICEA [11]. These two models have outperformed BRCAPRO and BCRAT in two previous validation studies [13,14]. However, we believe it is valuable to investigate the combination of BRCAPRO and BCRAT in order to determine (1) how much predictive value non-genetic risk factors add to BRCAPRO, (2) how much predictive value detailed family history adds to BCRAT, and (3) whether model combination can achieve competitive performance compared to developing a hybrid model from the ground up. Moreover, BRCAPRO and BCRAT continue to be widely used in clinical practice. A model that directly combines and improves upon predictions from BRCAPRO and BCRAT may be appealing to current users of these two models.

We consider two combination approaches: (1) penetrance modification and (2) ensemble learning. While we focus on BRCAPRO and BCRAT in this paper, these approaches can be applied in general to expand any Mendelian model to include additional risk factors. The first approach modifies the penetrance functions in BRCAPRO to account for the effects of the BCRAT risk factors. We develop a penetrance modification model, BRCAPRO + BCRAT (M), using a relative hazard approach that has similarities to the one used in IBIS.



**Figure 1.** Inputs to BRCAPRO and BCRAT and their overlap.

Ensemble learning consists of training multiple base models and combining their predictions. A wide variety of ensemble methods have been developed, including stacking [17], which involves training a meta-model to optimally combine predictions from the base models, bagging [18,19], which involves averaging models trained on bootstrap samples of the original data, and boosting [20], which involves constructing an ensemble by sequentially adding new models that are trained to correct the errors of previous ones. The rationale for these methods is that combining predictions can reduce variance and expand the set of functions that can be represented by any individual base model [21]. Extensive research, both empirical [21,22] and theoretical [23–27], has shown that ensembles can achieve performance gains over their base models, especially when the base models produce dissimilar predictions [28,29]. Debray et al. (2014) [30] demonstrated the value of aggregating published prediction models using real and simulated data on deep venous thrombosis and traumatic brain injury. Across various scenarios, model averaging and stacking outperformed model recalibration [31,32] and performed as well as or better than developing a new model from the ground up. Moreover, the authors noted that stacking is more efficient than model averaging, as stacking has fewer unknown parameters. In the setting of breast cancer risk prediction, Ming et al. (2019) [33] showed that boosting and random forests, a form of bagging, were able to achieve higher discriminatory accuracy than BCRAT and BOADICEA. In this paper, we develop a stacked logistic regression ensemble, BRCAPRO + BCRAT (E), that takes predictions from BRCAPRO and BCRAT as input.

We compare the performance of the combination models to the individual BRCAPRO and BCRAT models in simulations and a data application, where we use training data from the Newton-Wellesley Hospital (NWH) and validation data from the Cancer Genetics Network (CGN). In the data application, we use IBIS as a reference for comparison in order to evaluate the strategy of developing a hybrid model from the ground up versus the strategy of combining existing models.

## 2. Materials and Methods

### 2.1. Problem Definition

Given a female individual without a previous diagnosis of breast cancer who presents for risk assessment (the counselee), the goal is to predict her risk of developing invasive breast cancer within  $\tau$  years (where  $\tau$  is a pre-specified positive integer) based on family history and other risk factors while accounting for death from other causes as a competing risk. Below, we provide an overview of the risk prediction models compared in this paper. A more detailed description of the models with formal notation is provided in Appendix A.

## 2.2. Existing Models

### 2.2.1. BRCAPRO

BRCAPRO [8] estimates the probability of carrying a deleterious germline mutation in BRCA1 and BRCA2 using Bayes' rule, laws of Mendelian inheritance, mutation prevalence and penetrance, and family history of breast and ovarian cancer. In addition, it estimates future risk of breast and ovarian cancer based on the carrier probabilities and penetrances. The family history information used by BRCAPRO includes the following variables for each family member (missing values are allowed): gender, breast cancer status, age of onset of breast cancer if applicable, ovarian cancer status, age of onset of ovarian cancer if applicable, and current age or age at death. Information on preventative interventions (mastectomy/oophorectomy) and genetic testing results for BRCA1/BRCA2 can also be included.

BRCAPRO calculates the counselee's probability of having a given genotype using Bayes' rule and Mendelian laws of inheritance. This calculation depends on mutation prevalence and penetrance estimates obtained from published studies (see Appendix A for more details). The prevalences are ethnicity-specific (in particular, different prevalences are used for Ashkenazi Jewish and non-Ashkenazi Jewish families), and the penetrances, which are functions of age, are cancer- and sex-specific. The penetrance functions for non-carriers are based on rates from the Surveillance, Epidemiology, and End Results (SEER) program, and are race-specific, while the penetrance functions for carriers are from a meta-analysis of published studies [34]. After estimating the carrier probabilities, BRCAPRO calculates future risk of breast and ovarian cancer through a weighted average of the genotype-specific risks.

We ran BRCAPRO using the BayesMendel R package [35] version 2.1-6.1, selecting the crude risk option.

### 2.2.2. BCRAT

BCRAT [4–7] estimates the relative hazard of developing breast cancer based on age and the following risk factors: age at menarche, number of benign breast biopsies, age at first live birth, number of female first-degree relatives with breast cancer, and history of atypical hyperplasia. The relative hazard model includes interactions between age and number of biopsies as well as age at first live birth and number of affected relatives. The regression coefficients were estimated from U.S. case-control studies. Separate models were fit to data from White, African-American, Asian, and Hispanic women to obtain race/ethnicity-specific estimates. In the future risk calculation, the race-specific baseline hazard of breast cancer (based on SEER data) is multiplied by the relative hazard associated with the risk factors (see Appendix A).

We ran BCRAT using the BCRA R package (<https://cran.r-project.org/web/packages/BCRA/index.html>, accessed on 20 March 2020), version 2.1.

### 2.2.3. IBIS

In our data application, we examine the IBIS model [12,36] as a reference for comparison, as it combines detailed family history information with non-genetic risk factors. It first calculates carrier probabilities and risk of breast cancer based on family history, then incorporates additional risk factors (age at menarche, age at menopause, height, body mass index, age at first live birth, menopausal hormone therapy, atypical hyperplasia, lobular carcinoma in situ, breast density, and a polygenic risk score for breast cancer) via a relative hazard model. The carrier probabilities are calculated using a similar approach as in BRCAPRO; however, in addition to BRCA1 and BRCA2, IBIS considers a hypothetical low-penetrance susceptibility gene that acts as a surrogate for other unspecified breast cancer susceptibility genes. Estimates of the prevalence and penetrance of BRCA1 and BRCA2 were obtained from previously published studies, while the prevalence and penetrance of the hypothetical gene were estimated from a Swedish population-based study. The penetrance function for non-carriers is based on incidence rates from the Thames Cancer Registry.

IBIS calculates a weighted average of the cumulative penetrances for each genotype to obtain the risk of breast cancer conditional on family history only. In the final future risk calculation, IBIS multiplies the hazard of breast cancer conditional on family history by the relative hazard associated with the additional risk factors (see Appendix A for details).

Software for running IBIS is available at <http://www.ems-trials.org/riskevaluator/> (accessed on 20 April 2018). We used the command line program for version 8, selecting the competing mortality option.

### 2.3. Model Combination Approaches

#### 2.3.1. Penetrance Modification Model: BRCAPRO + BCRAT (M)

The penetrance modification approach combines BRCAPRO and BCRAT by using the BCRAT relative hazard to modify the calculation of the non-carrier future risk in BRCAPRO. Specifically, the baseline non-carrier hazard in the BRCAPRO future risk calculation is scaled by the BCRAT relative hazard based on a discrete proportional hazards model (see Appendix A for details). In the current implementation, we apply the modification to only the non-carrier future risk calculation, as BCRAT is not recommended for known carriers of BRCA1/2 mutations [37]. As in BRCAPRO, the final risk is a weighted average of the genotype-specific risks.

We refer to this model as the penetrance modification model, BRCAPRO + BCRAT (M), because the modification of the hazard function induces a modification of the corresponding penetrance function. This combination approach is similar to replacing the non-carrier future risk from BRCAPRO with the future risk from BCRAT. However, it is not identical because BRCAPRO and BCRAT use slightly different baseline hazards (see Figure A1 in Appendix B for plots of the hazards used in BCRAT and BRCAPRO).

The relative hazard approach for incorporating the BCRAT risk factors has similarities to the approach used by IBIS, except that IBIS averages the genotype-specific risks before incorporating non-genetic risk factors while BRCAPRO + BCRAT (M) incorporates the BCRAT risk factors before averaging the genotype-specific risks. The advantage of the latter approach is that it allows the effects of the BCRAT risk factors to differ by genotype. Genotype-specific effects have been observed for certain BCRAT risk factors, such as age at menarche (see [38] for a review). However, in general, the effects of the BCRAT risk factors on carriers are not well-studied; only a limited number of prospective studies have been done and they have had small sample sizes [38]. Therefore, the current version of BRCAPRO + BCRAT (M) modifies only the non-carrier hazard.

#### 2.3.2. Ensemble Model: BRCAPRO + BCRAT (E)

The second model combination approach involves training a stacked ensemble model [17] that uses BRCAPRO and BCRAT as the (pre-trained) base models. We consider a logistic regression ensemble that predicts  $\tau$ -year risk of breast cancer for fixed  $\tau$ , as well as a time-to-event ensemble that provides predictions for different time intervals.

As in Debray et al. (2014) [30], for a fixed value of  $\tau$  we can combine the  $\tau$ -year BRCAPRO and BCRAT predictions using a logistic regression model that includes the two predictions as covariates along with an interaction term (their product). Other covariates and/or published models can be included as inputs. We refer to this model as BRCAPRO + BCRAT (E).

Alternatively, we can use BRCAPRO and BCRAT predictions and their product as covariates in a time-to-event model. We consider a Fine-Gray proportional subdistribution hazards model for breast cancer, accounting for death as a competing risk (see Appendix A for details). The BRCAPRO and BCRAT predictions are calculated at baseline for a predefined time point  $\tau^*$  and do not vary with  $\tau^*$ . We refer to this model as BRCAPRO + BCRAT (E2).

In contrast to the penetrance modification model, which can be implemented using published parameter estimates, the ensemble models require training. If predictions from BRCAPRO and BCRAT are highly correlated in the training data, then multicollinearity can lead to unstable coefficient estimates [39]. Moreover, the ensemble models might perform

poorly in external validation if the training data are not representative of the validation data. Ideally, they should be trained using a prospective cohort that is representative of the target population. Otherwise, recalibration or reweighting methods can be used to account for differences in the covariate distributions between the training population and the target population. One widely used method is importance weighting [40], which weights each training observation by the ratio of the joint probability distributions of the covariates in the target and training populations [40]. The importance weights can be estimated using kernel mean matching [41], Kullback–Leibler importance estimation [42], or least squares importance fitting [43].

In the simulations and data application, we applied a square root transformation to the BRCAPRO and BCRAT predictions prior to fitting the ensemble models, as the distributions of the predictions were highly right-skewed. For BRCAPRO + BCRAT (E2), we used  $\tau^* = 5$  to compute the covariates.

#### 2.4. Model Evaluation Metrics

In the simulations and data application, we considered the binary outcome of being diagnosed with breast cancer within  $\tau = 5$  years. In the data application, we additionally considered the time-to-event outcome over the course of follow-up, as many of the counsees in the validation dataset were followed for more than five years; however, there was substantial variability in follow-up times.

We used five performance measures [44]: (1) the ratio of observed (O) to expected (E) events, where E is calculated by summing everyone's predicted probabilities (a measure of calibration, with 1 indicating perfect calibration); (2) the area under the receiver operating characteristic curve (AUC) or concordance (C) statistic, which is the probability that an individual who experiences the event has a higher score than an individual who does not, and is a measure of discrimination; (3) the Brier score, which is the mean squared difference between the predicted probabilities and actual outcomes; (4) the standardized net benefit (SNB) [45–47], which is the difference between the true positive rate and a weighted false positive rate, where the weight is based on a pre-specified threshold for classifying individuals as high-risk versus low-risk (the weight is the ratio of the odds of the threshold risk to the odds of the outcome); and (5) the logarithmic score [48], which is the negative log-likelihood. Calibration was assessed using both overall O/E and calibration plots of O/E by risk decile. We calculated SNB for only the binary outcome, using a 5-year risk threshold of 1.67% (the clinical 5-year risk threshold for eligibility for chemoprevention). We report the Brier score and logarithmic score in terms of relative difference with respect to BRCAPRO, as these metrics are prevalence-dependent, and as such are more difficult to interpret on their original scale.

In the data application, certain individuals in the validation dataset were censored before  $\tau = 5$  years. To account for censoring, we used inverse probability of censoring weights (IPCW) [49,50] to calculate the O/E, AUC, Brier score, and logarithmic score for the binary outcome; individuals with observed outcomes were used to calculate the performance measures and weighted by their inverse probability of not being censored by the minimum of (1) the end of the 5-year projection period and (2) the time at which they were diagnosed with breast cancer. Individuals who were censored were not directly used to calculate the performance measures, and were instead used to estimate the censoring distribution via the Kaplan–Meier estimator.

For the time-to-event outcome, O/E was calculated by comparing the numbers of observed and expected cases across the entire study period; for the expected number of cases E, we predicted risk up to the end of the individual follow-up time for each counselee, meaning that  $\tau$  varied across counsees. In addition, we used time-to-event versions of the C-statistic and logarithmic score. The time-to-event C-statistic [51] is the probability of an individual with a shorter time-to-event having a higher score than an individual with a longer time-to-event. A fixed prediction period of  $\tau = 10$  years was used to calculate the C-statistic (which requires the same prediction period for everyone), as ten years was the

maximum event time observed in the data. We used a version of the logarithmic score that accounted for competing risks [52]. In the absence of censoring, the logarithmic score is a strictly proper scoring rule for the problem of predicting a probability distribution [53,54]. We have previously shown that the competing risks version remains strictly proper under non-informative censoring [52].

We calculated 95% bootstrap confidence intervals (CIs) for all of the performance measures except for the time-to-event C-statistic, for which a 95% CI was obtained using perturbation resampling [51]. For each performance measure, we looked at pairwise comparisons of the models across bootstrap replicates of the validation set.

### 2.5. Simulations

We performed simulations to compare the 5-year performance of the combination models to that of the individual BRCAPRO and BCRAT models in a setting where the assumptions of the penetrance modification model hold.

First, we generated each counselee's baseline family history, consisting of (1) the family structure (i.e., the number of sisters, number of brothers, etc.), (2) dates of birth, (3) genotypes, (4) cancer ages, and (5) death ages.

We simulated pedigrees in order to mimic the family structures observed in real families from the CGN dataset (the validation dataset for the data application, described in Section 2.6.2). We restricted the family members to first- and second-degree relatives of the counselee.

For counselees, dates of birth and baseline dates for risk assessment were sampled from the CGN dataset. For non-counselees, dates of birth were generated relative to the counselee's date of birth by assuming that the age difference between a parent and a child had a mean of 27 and standard deviation of 6. We generated the birth dates of the counselee's parents and children based on the counselee's birth date, then the birth dates of the counselee's grandparents and siblings based on the birth dates of the parents, then the birth dates of the counselee's aunts and uncles based on the birth dates of the grandmothers.

Next, we generated the BRCA1/2 genotypes for each family member. We first generated the genotypes of the grandparents (founders) using the default Ashkenazi Jewish allele frequencies for BRCA1 and BRCA2 in BRCAPRO, which reflect a higher-risk population (CGN participants represent a higher-risk population than the general population, as they were selected for family history of cancer). For individuals in subsequent generations, we generated genotypes according to Mendelian inheritance.

For each individual, we generated baseline breast and ovarian cancer phenotypes conditional on genotype. Age of onset was sampled from  $\{1, 2, \dots, \text{baseline age}\}$ , with probabilities provided by the genotype-specific penetrance functions from BRCAPRO. The probability of being unaffected at baseline was provided by one minus the cumulative penetrance up to the baseline age. Counselees were assumed to be alive at baseline, while for each non-counselee we generated a death age from a distribution with mean of 80 and standard deviation of 15. If an individual had an age of onset greater than their age at death, then the individual's cancer status was changed to unaffected. Counselees with breast cancer at baseline were excluded from the analyses.

We then generated baseline BCRAT covariates (other than number of affected first-degree relatives and age at first live birth, which were calculated from the baseline family history) by sampling values from the distribution in the CGN. Values for different covariates were sampled independently of each other. The BCRAT covariates were used to modify the BRCAPRO non-carrier penetrance to obtain the BRCAPRO + BCRAT (M) non-carrier penetrance (Equation (A16)).

For counselees who did not have breast cancer at baseline, future ages of onset were generated from the BRCAPRO + BCRAT (M) penetrances (for carriers, the BRCAPRO + BCRAT (M) penetrances are the same as in BRCAPRO), which were rescaled to be conditional on not having developed cancer by the baseline age. Cases were defined as

counselees who developed breast cancer within five years of their baseline age. The 5-year outcomes were not subject to censoring.

We simulated 100,000 families in total. After excluding 4443 counselees with breast cancer at baseline, we used the first randomly generated 50,000 counselees to train the ensemble model (similar to the size of the training set in the data application; see Section 2.6.1) and the remaining 45,557 for validation. There were 814 cases in the training set and 724 cases in the validation set.

## 2.6. Data Application

Using data from Newton-Wellesley Hospital (NWH), we trained ensemble models for the binary outcome at  $\tau = 5$  (Model (A19)) and for the time-to-event outcome (Model (A20)). We validated these models, along with BRCAPRO + BCRAT (M), BRCAPRO, BCRAT, and IBIS, using data from the CGN. We assessed performance based on both the binary and time-to-event outcomes. In addition, we looked at performance stratified by family history, with strata defined based on the NCCN criteria for further genetic risk evaluation [55].

In the analyses, we excluded women with any of the following conditions prior to baseline: invasive breast cancer, ductal carcinoma in situ, lobular carcinoma in situ, bilateral mastectomy or bilateral oophorectomy. In addition, we excluded women who tested positive for BRCA1/2 prior to baseline (BCRAT requirement), women < 20 years old at baseline (BCRAT requirement), and women with a projection interval extending beyond 85 years of age (IBIS requirement).

The characteristics of the training and validation datasets are described below and summarized in Table 1.

**Table 1.** Main characteristics of the CGN and NWH cohorts.

Variable	Category	CGN	NWH
N		7314	37,881
Age (median [IQR])		47 [38, 57]	49 [43, 58]
Race (%)	White	6104 (83.5)	30,758 (81.2)
	Black	257 (3.5)	479 (1.3)
	Hispanic	694 (9.5)	548 (1.4)
	Asian	160 (2.2)	1228 (3.2)
	Native American	29 (0.4)	25 (0.1)
	Unknown	70 (1.0)	4843 (12.8)
Affected 1st-degree Relatives (%)	0	4171 (57.0)	32,197 (85.0)
	1	2496 (34.1)	5277 (13.9)
	2+	647 (8.8)	407 (1.1)
Follow-up Time (median [IQR])		7.3 [6.0, 8.3]	6.7 [6.3, 7.2]
Censored < 5 Years (%)		934 (12.8)	0 (0.0)
Cases (%)		159 (2.2)	714 (1.9)
5-year Cases (%)		112 (1.5)	495 (1.3)

### 2.6.1. Training Dataset (NWH)

After applying the exclusion criteria, the training cohort consisted of 37,881 women who visited the breast imaging department of NWH in Newton, Massachusetts for screening or diagnostic imaging from February 2007 through December 2009. During the initial (baseline) visit, information was collected on personal and family history of cancer, reproductive history, sociodemographic factors, and lifestyle factors. Family history was limited to relatives with cancer. Breast cancer diagnoses through 2015 were determined from the Massachusetts State Cancer Registry, Partners Hospital Cancer Registries, and patient self-reporting. The median age of the counselees was 49, with an interquartile range (IQR) of 43–58; 30,758 (81.2%) of the counselees were White, while 5684 (15.0%) had at least

one affected first- or second-degree relative. The median follow-up time was 6.7 years (IQR 6.3–7.2). All counsees were followed for at least six years. There were 495 counsees (1.3%) who developed breast cancer within five years of baseline and 714 counsees (1.9%) who developed breast cancer over the course of follow-up.

Because the NWH cohort represents a general screening population and the CGN validation cohort (described below) represents a higher-risk population enriched for family history of cancer, we applied importance weights to the training data based on the distributions of the BCRAT covariates, 5-year BCRAT predictions, and 5-year BRCAPRO predictions when we fit the ensemble models. The weights were estimated using least squares importance fitting via the *densratio* R package (<https://cran.r-project.org/web/packages/densratio/index.html> (accessed on 20 April 2018)).

### 2.6.2. Validation Dataset (CGN)

The validation cohort consisted of 7314 women who enrolled in the CGN, a national research network of fifteen academic medical centers that was established for the purpose of studying inherited predisposition to cancer. Enrollment began in 1999 and ended in 2010. One of the criteria for enrollment was a personal and/or family history of cancer. Participants provided information on personal and family history of cancer, sociodemographic factors, and lifestyle factors through an initial (baseline) phone interview and annual follow-up updates. From 2009 onward, information was collected on reproductive history, cancer treatments, cancer screening results, and genetic testing results.

The median age of the counsees was 47 (IQR 38–57); 6104 (83.5%) of the counsees were White, while 3143 (42.9%) had at least one female first-degree relative with breast cancer. The median follow-up time was 7.3 years (IQR 6.0–8.3) and 934 (12.8%) counsees were censored within five years of baseline without being diagnosed with breast cancer. Of the counsees, 159 (2.2%) developed breast cancer during follow-up, with 112 of the diagnoses occurring within five years of baseline. Demographic characteristics stratified by center are provided in Table A4 of Appendix D. Because follow-up times and breast cancer incidence rates varied by center, we estimated the censoring distribution (which was needed to calculate several of the performance measures) by separately fitting a Kaplan–Meier curve for each center.

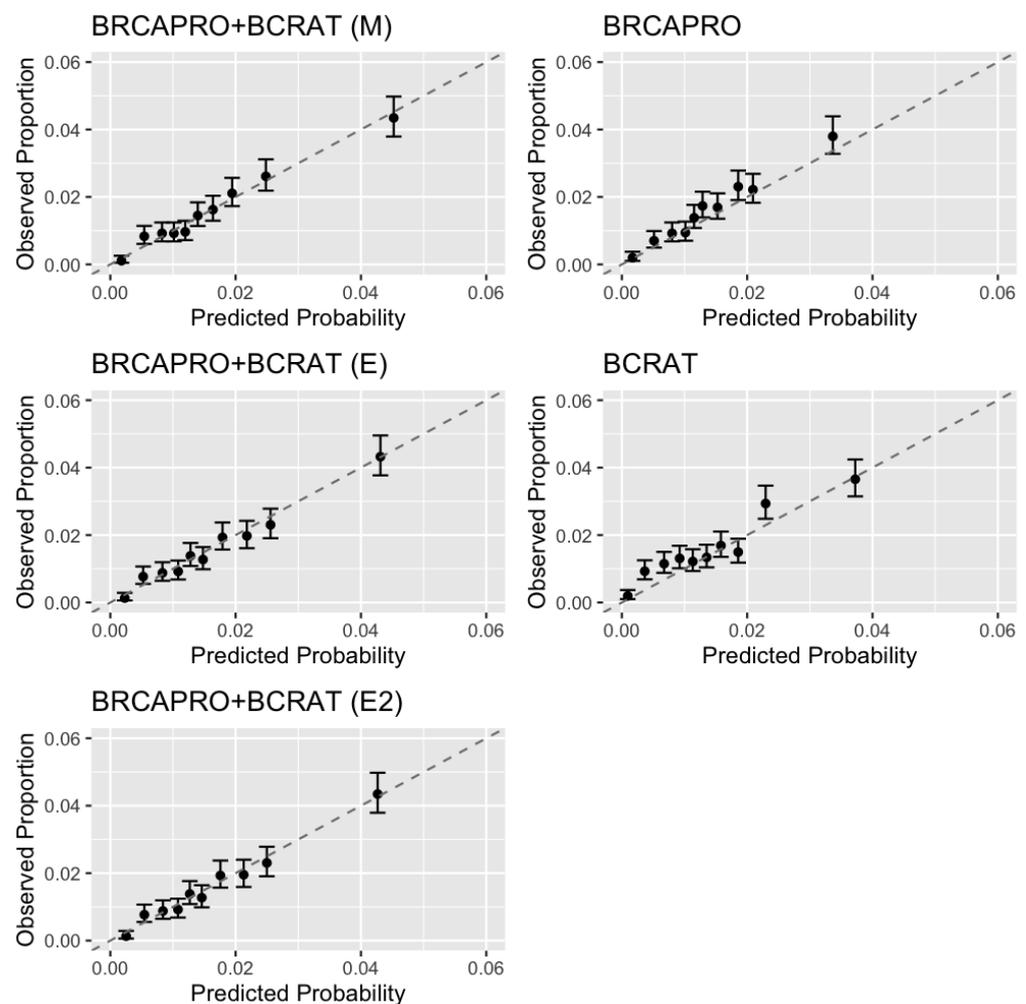
Information on certain risk factors was missing or incomplete. We did not have information on atypical hyperplasia (used in BCRAT and IBIS), breast density (used in IBIS), polygenic risk scores (used in IBIS), or hormone replacement therapy (used in IBIS). Participants were asked whether they had ever had a benign breast biopsy, but were not asked about the number of biopsies (categorized as 0, 1, or  $\geq 2$  in BCRAT). Because participants were asked about reproductive history starting in 2009, 4157 (56.8%) were missing age at menarche (used in BCRAT and IBIS). Information on Ashkenazi Jewish heritage (used in BRCAPRO and IBIS) was not available for the University of Washington (UWASH) center. We coded the missing variables according to the specifications of the software for each model. Number of breast biopsies was coded as 1 for participants who indicated that they had previously had a biopsy.

## 3. Results

### 3.1. Simulation Results

The performance measures are shown in Table 2 and calibration plots are shown in Figure 2. We chose not to report the performance of the IBIS model in this section, as the data generating model is more favorable to BRCAPRO and BCRAT. BRCAPRO + BCRAT (M), the true model, had the best performance, although the ensemble models were able to achieve similar performance to the true model and performance gains over BRCAPRO and BCRAT. The combination models were well-calibrated overall, with O/E = 1.01 (95% CI 0.94–1.08) for BRCAPRO + BCRAT (M), O/E = 0.98 (95% CI 0.91–1.04) for BRCAPRO + BCRAT (E), and O/E = 0.99 (95% CI 0.92–1.05) for BRCAPRO + BCRAT (E2). BRCAPRO and BCRAT underpredicted the number of cases, with O/E = 1.15 (95% CI 1.07–1.23) for BRCAPRO and O/E = 1.14 (95% CI 1.05–1.21) for BCRAT. The combination models were

well-calibrated in each decile of risk (Figure 2), while BRCAPRO and BCRAT were more prone to underpredicting risk in certain deciles. The combination models had slightly higher AUCs than BRCAPRO and BCRAT: 0.69 (95% CI 0.67–0.71 for BRCAPRO + BCRAT (M), 0.68 (95% CI 0.67–0.70) for BRCAPRO + BCRAT (E) and BRCAPRO + BCRAT (E2), 0.67 (95% CI 0.65–0.69) for BRCAPRO, and 0.66 (95% CI 0.64–0.68) for BCRAT. In addition, the combination models performed better than BRCAPRO and BCRAT with respect to the Brier score, logarithmic score, and SNB. Across 1000 bootstrap replicates of the validation dataset, BRCAPRO + BCRAT (M) outperformed BRCAPRO and BCRAT with respect to all performance measures in more than 95% of the replicates and the ensemble models outperformed BRCAPRO and BCRAT with respect to all performance measures except O/E in more than 95% of the replicates. The ensemble models had better O/E ratios than BRCAPRO and BCRAT more than 92% of the time. BRCAPRO + BCRAT (M) outperformed BRCAPRO + BCRAT (E) and BRCAPRO + BCRAT (E2) more than 97% of the time with respect to each of AUC, Brier score, and logarithmic score.



**Figure 2.** Calibration plots by decile of risk for five-year predictions in a simulated dataset with 45,557 counselees (724 cases). For each model, we grouped individuals by decile of risk and plotted the observed proportion of women who developed cancer (with 95% Wilson CI) versus the predicted probability (sum of risk predictions) within each decile.

**Table 2.** Five-year performance on a simulated dataset with 45,557 probands (717 cases). B+B: BRCAPRO + BCRAT. ΔBS: % relative improvement in Brier Score compared to BRCAPRO. ΔLS: % relative improvement in logarithmic score compared to BRCAPRO. The “Comparisons Across Bootstrap Replicates” section shows pairwise comparisons involving the combination models across 1000 bootstrap replicates of the validation dataset; the row for  $A > B$  shows the proportion of bootstrap replicates where model A outperformed model B with respect to each metric. Proportions  $> 0.5$  are highlighted in blue (with darker shades of blue for higher proportions) and proportions  $\leq 0.5$  are highlighted in red (with darker shades of red for lower proportions).

	O/E	AUC	SNB	ΔBS	ΔLS
<b>Performance Metrics</b>					
B+B (M)	1.01 (0.94, 1.08)	0.69 (0.67, 0.71)	0.26 (0.21, 0.30)	0.25 (0.09, 0.41)	1.12 (0.60, 1.66)
B+B (E)	0.98 (0.91, 1.04)	0.68 (0.67, 0.70)	0.25 (0.20, 0.29)	0.12 (−0.02, 0.25)	0.63 (0.18, 1.05)
B+B (E2)	0.99 (0.92, 1.05)	0.68 (0.67, 0.70)	0.24 (0.19, 0.29)	0.16 (0.01, 0.29)	0.65 (0.21, 1.07)
BRCAPRO	1.15 (1.07, 1.23)	0.67 (0.65, 0.69)	0.21 (0.17, 0.25)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
BCRAT	1.14 (1.05, 1.21)	0.66 (0.64, 0.68)	0.20 (0.15, 0.24)	−0.21 (−0.46, 0.06)	−1.15 (−2.10, −0.16)
<b>Comparisons Across Bootstrap Replicates</b>					
B+B(M)>B+B(E)	0.570	0.994	0.725	0.999	1.000
B+B(M)>B+B(E2)	0.504	0.995	0.795	0.972	1.000
B+B(M)>BRCAPRO	0.978	0.999	0.998	0.998	1.000
B+B(M)>BCRAT	0.962	1.000	1.000	1.000	1.000
B+B(E)>B+B(E2)	0.312	0.919	0.808	0.011	0.133
B+B(E)>BRCAPRO	0.944	0.999	1.000	0.961	0.998
B+B(E)>BCRAT	0.924	1.000	0.999	0.994	1.000
B+B(E2)>BRCAPRO	0.953	0.999	0.995	0.979	0.999
B+B(E2)>BCRAT	0.938	1.000	0.998	0.996	1.000

### 3.2. Data Application Results

#### 3.2.1. Five-Year Binary Outcome

The performance measures based on the five-year outcome are shown in Table 3 (overall performance) and Table 4 (performance stratified by family history). Figure 3 shows the distributions of the predictions as well as pairwise correlations between models, while Figure 4 shows the calibration plots. The weights from the ensemble models are provided in Appendix C.

As seen in Figure 3, the lowest correlations were observed between models with smaller overlaps of the input variables, as expected. BRCAPRO and BCRAT had the lowest correlation ( $\rho = 0.44$ ). Even models with higher correlations displayed divergent predictions. Correlations of predictions from BRCAPRO + BCRAT (M) with predictions from each of the other models ranged from  $\rho = 0.78$  with BRCAPRO to  $\rho = 0.93$  with BRCAPRO + BCRAT (E). BRCAPRO + BCRAT (E) and BRCAPRO + BCRAT (E2), which assigned a higher weight to BCRAT than to BRCAPRO (Appendix C), were more highly correlated with BCRAT ( $\rho = 0.93$ ) than with BRCAPRO ( $\rho = 0.67$  for (E) and  $\rho = 0.62$  for (E2)).

On aggregate, BRCAPRO + BCRAT (M) (O/E = 1.03, 95% CI 0.84–1.23) and IBIS (O/E = 0.98, 95% CI 0.80–1.17) were well-calibrated, while BCRAT (O/E = 1.15, 95% CI 0.94–1.37), BRCAPRO + BCRAT (E) (O/E = 1.17, 95% CI 0.96–1.40), BRCAPRO + BCRAT (E2) (O/E = 1.20, 95% CI 0.98–1.43), and BRCAPRO (O/E = 1.30, 95% CI 1.07–1.56) underestimated risk. When calibration was considered by decile (Figure 4), 8 out of 60 intervals overall failed to cross the diagonal, compared to the result of 3 expected by chance when all models are well-calibrated. BRCAPRO + BCRAT (M) overestimated risk in the top decile of risk and IBIS overestimated risk in the second highest decile, while BRCAPRO and BCRAT underestimated risk in other deciles. All models except BRCAPRO overpredicted risk by a certain amount in the second-smallest decile, suggesting that the source of this pattern may reside in specific characteristics of the distribution of covariates other than family history.

The AUCs were 0.68 (95% CI 0.63–0.72) for the combination models, 0.67 for IBIS (95% CI 0.62–0.71), 0.66 for BCRAT (95% CI 0.61–0.71), and 0.65 (95% CI 0.60–0.69) for BRCAPRO. IBIS had the highest SNB (SNB = 0.28, 95% CI 0.16–0.38), followed by BRCAPRO + BCRAT (M) (SNB = 0.24, 95% CI 0.13–0.35) and the ensemble models (SNP = 0.24, 95% CI 0.12–0.34).

for (E2), SNB = 0.23 and SNP = 0.23, 95% CI 0.10–0.33 for (E)). All models performed similarly with respect to the Brier score and logarithmic score. Across 1000 bootstrap replicates, BRCAPRO + BCRAT (M) outperformed BRCAPRO and BCRAT with respect to all performance measures except the Brier score in the majority of the replicates. Each ensemble model outperformed BRCAPRO and BCRAT with respect to all metrics except O/E in the majority of the replicates. Furthermore, each of the three combination models outperformed IBIS with respect to AUC, Brier score, and logarithmic score in the majority of the replicates, although IBIS had better calibration and SNB in most replicates.

Among counselees who met the NCCN criteria for further genetic risk evaluation (Table 4), the combination models and IBIS had higher AUCs and SNBs than BRCAPRO and BCRAT. BRCAPRO + BCRAT (M) and IBIS overestimated risk while all other models except BCRAT underestimated risk. Among counselees who did not meet the NCCN criteria, all models underestimated risk. BRCAPRO had a slightly lower AUC than the other models, and IBIS had the highest SNB.

**Table 3.** Five-year performance in the entire CGN cohort. B+B: BRCAPRO + BCRAT. ΔBS: % relative improvement in Brier Score compared to BRCAPRO. ΔLS: % relative improvement in logarithmic score compared to BRCAPRO. The “Comparisons Across Bootstrap Replicates” section shows pairwise comparisons involving the combination models across 1000 bootstrap replicates of the validation dataset; the row for  $A > B$  shows the proportion of bootstrap replicates where model A outperformed model B with respect to each metric. Proportions > 0.5 are highlighted in blue (with darker shades of blue for higher proportions) and proportions ≤ 0.5 are highlighted in red (with darker shades of red for lower proportions).

	O/E	AUC	SNB	ΔBS	ΔLS
<b>Performance Metrics</b>					
B+B (M)	1.03 (0.84, 1.23)	0.68 (0.63, 0.72)	0.24 (0.13, 0.35)	0.21 (−0.42, 0.84)	1.61 (−0.49, 3.59)
B+B (E)	1.17 (0.96, 1.40)	0.68 (0.63, 0.72)	0.23 (0.10, 0.33)	0.38 (−0.10, 0.84)	1.74 (−0.06, 3.41)
B+B (E2)	1.20 (0.98, 1.43)	0.68 (0.63, 0.72)	0.24 (0.12, 0.34)	0.40 (−0.07, 0.85)	1.81 (0.06, 3.48)
BRCAPRO	1.30 (1.07, 1.56)	0.65 (0.60, 0.69)	0.17 (0.05, 0.25)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
BCRAT	1.15 (0.94, 1.37)	0.66 (0.61, 0.71)	0.18 (0.07, 0.28)	0.22 (−0.51, 0.85)	0.88 (−2.03, 3.60)
IBIS	0.98 (0.80, 1.17)	0.67 (0.62, 0.71)	0.28 (0.16, 0.38)	−0.05 (−0.69, 0.52)	0.89 (−1.43, 3.07)
<b>Comparisons Across Bootstrap Replicates</b>					
B+B(M)>B+B(E)	0.841	0.421	0.702	0.208	0.412
B+B(M)>B+B(E2)	0.874	0.382	0.534	0.225	0.381
B+B(M)>BRCAPRO	0.943	0.920	0.933	0.727	0.934
B+B(M)>BCRAT	0.811	0.763	0.982	0.503	0.783
B+B(M)>IBIS	0.475	0.709	0.186	0.823	0.774
B+B(E)>B+B(E2)	0.963	0.310	0.001	0.288	0.266
B+B(E)>BRCAPRO	0.983	0.965	0.913	0.936	0.970
B+B(E)>BCRAT	0.066	0.844	0.870	0.895	0.907
B+B(E)>IBIS	0.214	0.717	0.119	0.919	0.806
B+B(E2)>BRCAPRO	0.987	0.955	0.950	0.953	0.980
B+B(E2)>BCRAT	0.047	0.870	0.937	0.895	0.938
B+B(E2)>IBIS	0.181	0.739	0.189	0.918	0.823

### 3.2.2. Time-to-Event Outcome

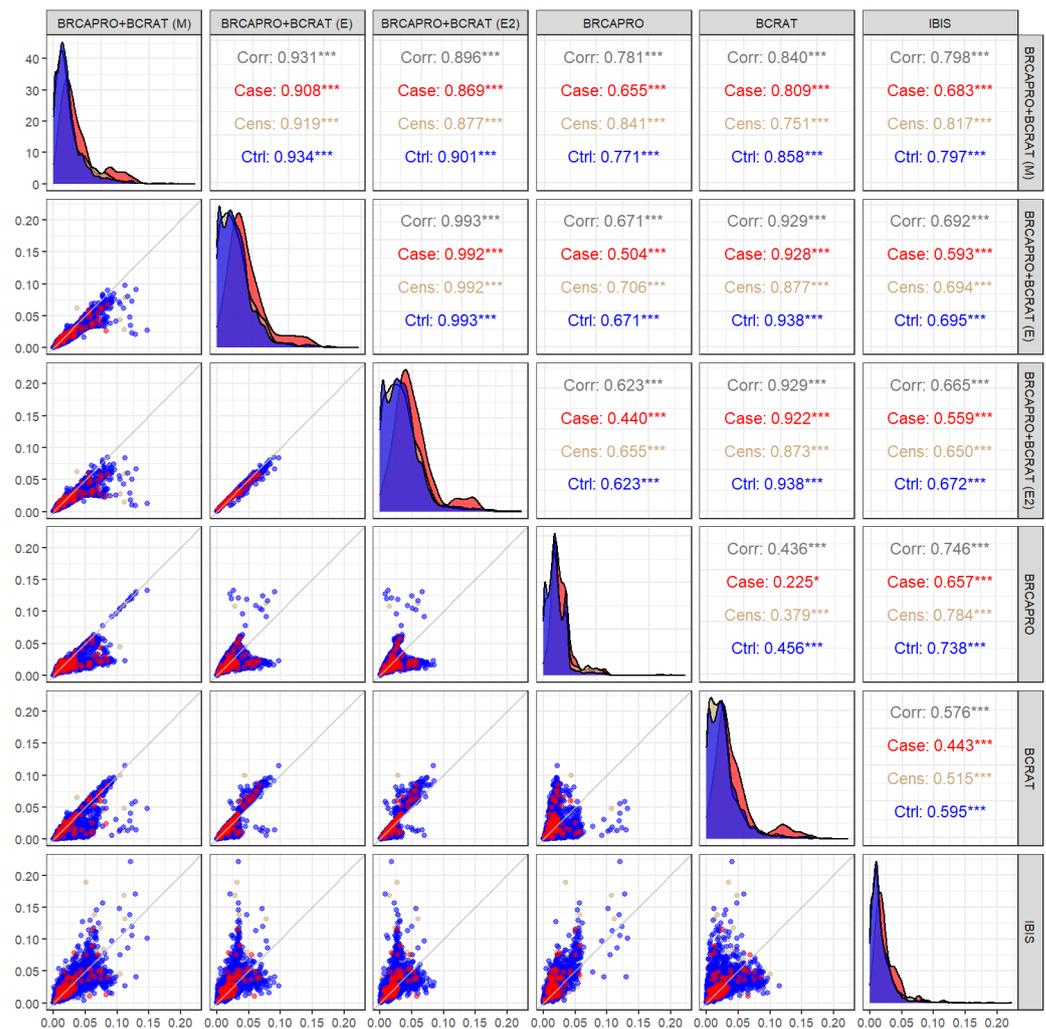
The performance measures based on the time-to-event outcome are shown in Table A5 (overall performance) and Table A6 (performance stratified by family history) in Appendix D. BRCAPRO + BCRAT (E) was excluded from the analyses because it only provides five-year risks. We did not calculate the logarithmic score for BRCAPRO + BCRAT (E2) or IBIS because the cause-specific distribution for competing mortality, which is needed to calculate the likelihood, is not explicitly modelled by BRCAPRO + BCRAT (E2) and is not available as an output from the IBIS software.

Overall, the relative performance of the models for the time-to-event outcome was similar to the relative performance for the five-year outcome; however, among counselees meeting the NCCN criteria BCRAT had worse discriminatory accuracy for the time-to-event outcome than the other models. O/E and discriminatory accuracy did not change

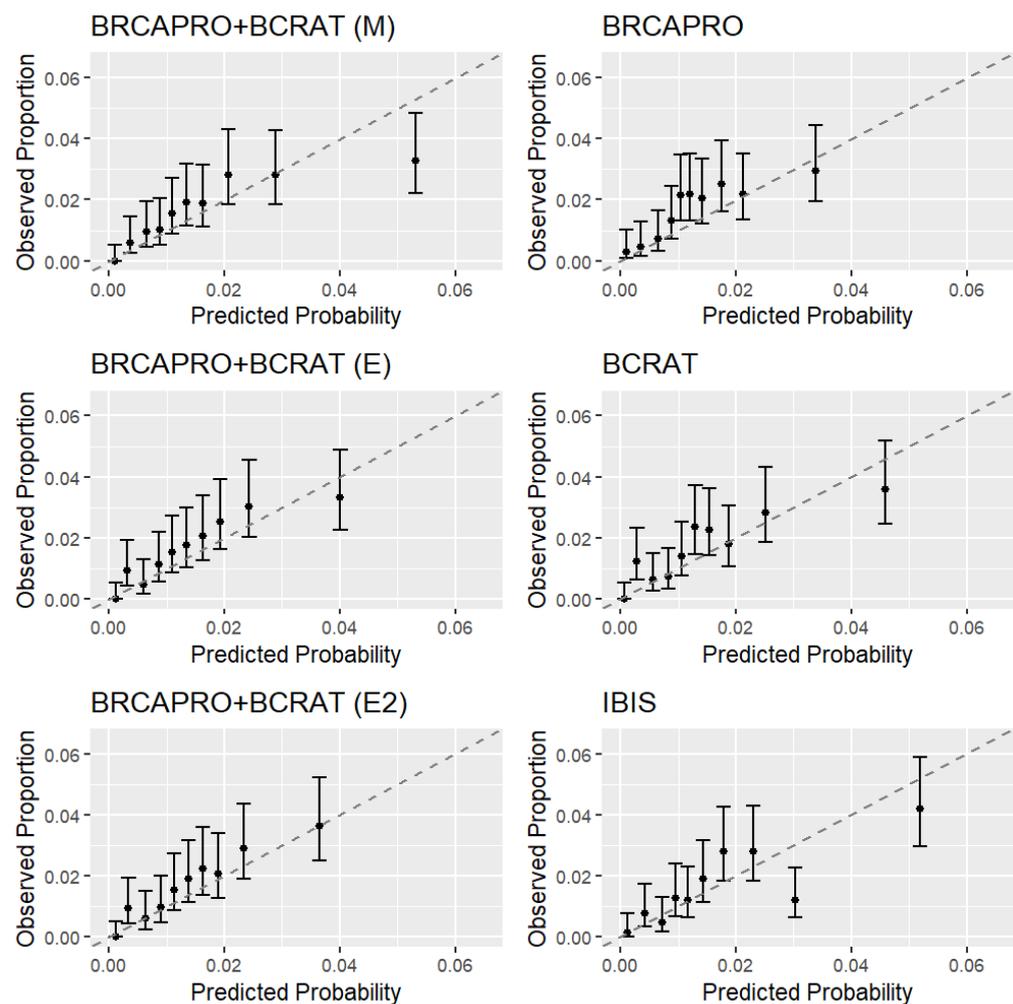
substantially compared to Tables 3 and 4, and differences in Brier score and logarithmic score across models were small. BRCAPRO + BCRAT (M) and BRCAPRO + BCRAT (E2) performed similarly to BCRAT and IBIS overall (Table A5), and had higher C-statistics than BCRAT in the subset of counselees meeting the NCCN criteria (Table A6). The combination models had similar or better calibration and discrimination compared to BRCAPRO overall as well as within each family history stratum.

**Table 4.** Five-year performance in the CGN cohort stratified by family history (whether or not the proband met the NCCN criteria for further genetic risk evaluation [55]; in applying the criteria, we only used information on breast and ovarian cancer diagnoses in relatives). B+B: BRCAPRO + BCRAT. ΔBS: % relative improvement in Brier Score compared to BRCAPRO. The “Comparisons Across Bootstrap Replicates” section shows pairwise comparisons involving the combination models across 1000 bootstrap replicates of the validation dataset; the row for  $A > B$  shows the proportion of bootstrap replicates where model A outperformed model B with respect to each metric. Proportions  $> 0.5$  are highlighted in blue (with darker shades of blue for higher proportions) and proportions  $\leq 0.5$  are highlighted in red (with darker shades of red for lower proportions).

	O/E	AUC	SNB	ΔBS	ΔLS
<b>Strong Family History (34 cases)</b>					
<b>Performance Metrics</b>					
B+B (M)	0.81 (0.55, 1.09)	0.71 (0.63, 0.79)	0.44 (0.18, 0.59)	0.75 (−1.20, 2.14)	3.62 (−2.27, 8.06)
B+B (E)	1.07 (0.74, 1.44)	0.71 (0.62, 0.79)	0.41 (0.17, 0.57)	1.19 (−0.16, 2.54)	4.15 (−0.77, 8.39)
B+B (E2)	1.14 (0.78, 1.53)	0.71 (0.62, 0.79)	0.41 (0.17, 0.57)	1.11 (−0.19, 2.49)	3.90 (−1.07, 8.29)
BRCAPRO	1.32 (0.91, 1.79)	0.66 (0.58, 0.74)	0.30 (0.08, 0.47)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
BCRAT	1.03 (0.71, 1.39)	0.66 (0.56, 0.75)	0.31 (0.06, 0.47)	0.77 (−1.36, 2.65)	1.55 (−6.64, 8.24)
IBIS	0.74 (0.50, 1.00)	0.69 (0.60, 0.77)	0.41 (0.14, 0.57)	−0.15 (−2.52, 1.46)	1.12 (−5.56, 5.92)
<b>Comparisons Across Bootstrap Replicates</b>					
B+B(M)>B+B(E)	0.339	0.593	0.634	0.192	0.318
B+B(M)>B+B(E2)	0.409	0.584	0.633	0.272	0.408
B+B(M)>BRCAPRO	0.615	0.931	0.917	0.794	0.892
B+B(M)>BCRAT	0.278	0.906	0.965	0.463	0.778
B+B(M)>IBIS	0.951	0.706	0.718	0.880	0.920
B+B(E)>B+B(E2)	0.704	0.521	0.423	0.699	0.718
B+B(E)>BRCAPRO	0.841	0.853	0.905	0.956	0.949
B+B(E)>BCRAT	0.410	0.945	0.879	0.856	0.941
B+B(E)>IBIS	0.751	0.669	0.532	0.919	0.897
B+B(E2)>BRCAPRO	0.876	0.834	0.903	0.949	0.933
B+B(E2)>BCRAT	0.335	0.954	0.877	0.772	0.915
B+B(E2)>IBIS	0.664	0.652	0.541	0.901	0.874
<b>Less Family History (78 cases)</b>					
<b>Performance Metrics</b>					
B+B (M)	1.16 (0.94, 1.44)	0.65 (0.60, 0.71)	0.16 (0.02, 0.28)	−0.02 (−0.41, 0.37)	0.80 (−0.92, 2.47)
B+B (E)	1.21 (0.98, 1.49)	0.66 (0.61, 0.71)	0.14 (0.01, 0.27)	0.03 (−0.22, 0.26)	0.80 (−0.54, 2.08)
B+B (E2)	1.22 (0.98, 1.50)	0.66 (0.61, 0.71)	0.16 (0.03, 0.29)	0.10 (−0.16, 0.33)	1.02 (−0.29, 2.34)
BRCAPRO	1.28 (1.04, 1.59)	0.63 (0.58, 0.68)	0.10 (−0.02, 0.21)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
BCRAT	1.20 (0.97, 1.48)	0.66 (0.61, 0.71)	0.12 (−0.00, 0.24)	−0.01 (−0.41, 0.35)	0.68 (−1.55, 2.92)
IBIS	1.14 (0.92, 1.41)	0.66 (0.61, 0.71)	0.22 (0.09, 0.34)	−0.01 (−0.38, 0.39)	0.81 (−1.38, 3.11)
<b>Comparisons Across Bootstrap Replicates</b>					
B+B(M)>B+B(E)	0.937	0.275	0.623	0.333	0.490
B+B(M)>B+B(E2)	0.940	0.196	0.429	0.233	0.329
B+B(M)>BRCAPRO	0.964	0.875	0.822	0.432	0.813
B+B(M)>BCRAT	0.934	0.223	0.900	0.424	0.542
B+B(M)>IBIS	0.101	0.396	0.121	0.429	0.476
B+B(E)>B+B(E2)	0.961	0.126	0.000	0.001	0.008
B+B(E)>BRCAPRO	0.979	0.958	0.806	0.630	0.890
B+B(E)>BCRAT	0.055	0.385	0.687	0.715	0.588
B+B(E)>IBIS	0.075	0.498	0.106	0.594	0.477
B+B(E2)>BRCAPRO	0.982	0.955	0.875	0.821	0.923
B+B(E2)>BCRAT	0.043	0.446	0.875	0.939	0.742
B+B(E2)>IBIS	0.069	0.532	0.168	0.744	0.579



**Figure 3.** Scatter plots, density plots, and pairwise correlations of five-year predictions in the CGN data. Red corresponds to data from cases, blue corresponds to data from controls (individuals who did not develop breast cancer within five years), and beige corresponds to data from counselees censored before five years. Corr: Pearson correlation, cens: counselees censored before five years, ctrl: controls. The lower diagonal panels show scatter plots of the predictions from each pair of models. For example, the panel in the second row, first column has predictions from BRCAPRO + BCRAT (E) on the x-axis and predictions from BRCAPRO + BCRAT (M) on the y-axis. The diagonal panels show density plots of the predictions from each model, stratified by case-control status. For example, the panel in the first row, first column shows the distribution of predictions from BRCAPRO + BCRAT (M). The upper diagonal panels show the Pearson correlations between predictions from each pair of models. For example, the first row, second column shows the overall correlation, correlation among cases, correlation among censored counselees, and correlation among controls for predictions from BRCAPRO + BCRAT (M) and BRCAPRO + BCRAT (E). \*\*\* =  $p < 0.001$ .



**Figure 4.** Calibration plots by decile of risk for five-year predictions in CGN. For each model, we grouped individuals by decile of risk and plotted the observed proportion of women who developed cancer (with 95% Wilson CI) versus predicted probability (sum of risk predictions) within each decile. In computing the observed proportions, the inverse probabilities of the censoring weights were used to account for censoring.

#### 4. Discussion

Model combination is a way to systematically integrate information from different models to achieve more comprehensive and accurate risk assessment. By leveraging existing models instead of building new ones from the ground up, it is possible to save considerable time and effort, remove barriers to adoption if the ingredient models are well-accepted, and, in the case of ensemble approaches such as stacking, permit reuse of proprietary models as long as predictions can be obtained even when the full model is not available.

We used model combination to expand a widely used Mendelian model, BRCAPRO, to include non-genetic risk factors from a widely used non-Mendelian model. We considered two approaches, namely, penetrance modification and ensemble learning. The penetrance modification model BRCAPRO + BCRAT (M) and the ensemble models BRCAPRO + BCRAT (E) for binary outcomes and BRCAPRO + BCRAT (E2) for time-to-event outcomes all achieved accuracy gains over BRCAPRO and BCRAT in simulations and data from the CGN, showing the value of model combination. Augmenting the family history input to BRCAPRO with non-genetic risk factors and augmenting the BCRAT risk factors with more detailed family history information both led to improvements, and could provide a better

tool to users of either BRCAPRO or BCRAT as well as to clinicians who use both and then qualitatively integrate the two estimates for their patients.

In data simulated under the penetrance modification model, all combination models outperformed BRCAPRO and BCRAT with respect to calibration, discrimination, net benefit, and prediction accuracy based on the Brier score and logarithmic score. In the CGN cohort, we additionally validated IBIS, an existing model that combines detailed family history with non-genetic risk factors. The penetrance modification model achieved comparable performance to IBIS overall, outperforming BRCAPRO with respect to each performance measure and outperforming BCRAT with respect to each measure except the Brier score in a large majority of bootstrap replicates. The ensemble models outperformed BRCAPRO with respect to each measure and BCRAT with respect each measure other than calibration, showing worse overall calibration than the penetrance modification model. In the subset of women with a strong family history of breast/ovarian cancer based on NCCN criteria, all combination models showed notable improvements in discrimination and net benefit over BRCAPRO and BCRAT.

While the AUC gains of 2–3% achieved by the combination models over BRCAPRO and BCRAT in the entire CGN cohort may not change the clinical implications for most individuals, the net benefit measure, which aims to quantify the expected benefits versus harms of treatment and has been used to assess the clinical utility of numerous models in cancer research and other fields [45,46], provides additional information on the value of the combination models. The 5–7% gain in standardized net benefit (based on a high-risk threshold of 1.67%, the FDA threshold for chemoprevention eligibility) indicates that there is a subset of individuals who would be accurately classified by the combination models and inaccurately classified by BRCAPRO/BCRAT. The combination models would be highly valuable for this subgroup. In particular, among women with a strong family history of cancer, the combination models achieved a 5% improvement in AUC and 10–14% improvement in standardized net benefit over BRCAPRO and BCRAT. In this subgroup, early screening and prevention measures can substantially reduce cancer risk and mortality [56]. Figure 3 highlights the substantial number of counselees for whom differences in model predictions are large enough to warrant different discussions about the expected benefits of preventative options.

BCRAT performed well in the entire CGN cohort, although it had less discrimination and net benefit than the combination models and IBIS among women with a strong family history, highlighting the importance of collecting detailed family history information for higher-risk subgroups [37]. Furthermore, BCRAT is not designed to calculate risk for known BRCA1/2 carriers, while the other four models all take into account genetic testing results (which the ensemble models do indirectly through BRCAPRO risk prediction).

The additional inputs in the combination models compared to BRCAPRO or BCRAT alone require additional data collection effort; however, models integrating detailed family history and non-genetic risk factors are already used in clinical practice (IBIS and BOADICEA), and these models use similar inputs compared to our combination models. Furthermore, missing values are allowed in our combination models, which can help to reduce the burden of data collection in clinical practice.

A limitation of our study arises from missing information on BCRAT and IBIS risk factors in the CGN dataset (atypical hyperplasia, age at menarche, and for IBIS, breast density, hormone replacement therapy, and polygenic risk scores). This could potentially have affected the discrimination of BCRAT, IBIS, and the combination models as well as their relative ranking. To mitigate this concern, we observe that the models retain relatively good discrimination that is comparable to previous studies [13,16]. Furthermore, the CGN did not collect genetic testing information for non-counselees, which could considerably improve the discrimination of BRCAPRO, IBIS, and the combination models [57].

Another limitation is that the CGN and NWH cohorts are predominantly non-Hispanic white (>80%), with <5% Black and <10% Hispanic. The combination models evaluated here do not overcome the potential limitations of BRCAPRO and BCRAT in predicting risk

among racial/ethnic minorities, at least as evaluated in the current datasets. While one recent study evaluated BRCAPRO, BCRAT, IBIS, and BRCAPRO + BCRAT (M) in a diverse US cohort and found no significant difference in model performance between Black and White women [58], BCRAT has shown only modest discrimination in other validation studies focusing on African-American, Asian-American, and US-born Hispanic women [5–7], and its performance has varied in non-US populations [59]. Moreover, other validation studies of BRCAPRO in minority population have shown its clinical utility for predicting carrier probabilities [60,61], although its performance for predicting future risk has not been evaluated. Future work should evaluate the combination models among more diverse populations.

A further challenge is that the NWH and CGN cohorts are not representative of the same clinical population. The NWH cohort is a lower-risk cohort; as seen in Table 1, it has a lower proportion of women with a first-degree family history of breast cancer. In addition, the family history information available for the NWH cohort is less detailed than that for the CGN cohort. While this allows us to assess the robustness of a model to heterogeneity across different datasets, it creates a challenge for the ensemble models, which are trained in NWH and validated in CGN, compared to the penetrance modification model, which leverages literature estimates based on populations potentially more similar to CGN. We used importance weighting to address this limitation; however, this approach remains an approximation and relies on accurate estimation of the probability distributions of risk factors in the training and target populations. The performance of the ensemble approach could potentially be improved by training on data that are more representative of the validation data. This is supported by the fact that the ensemble models were well-calibrated in the simulations, where the training and validation datasets were both generated under BRCAPRO + BCRAT (M).

The two combination approaches considered here each have their strengths and limitations. Ensembling via logistic regression calibrates the model to the training data. The penetrance modification model, on the other hand, relies on previously published estimates of prevalence, penetrance, and relative hazards. These features can be a strength or a limitation depending on how well the training dataset represents the target population. While ensembling can sometimes be successful without training weights (for example, predictions can be combined using a simple average), in our approach we used stacking to train weights on individual-level data. This opens opportunities for recalibration to improve suitability for a population different from the training population.

A disadvantage of estimating ensemble weights via logistic regression is the need to estimate the censoring distribution when there are censees in the training dataset who are censored before the last time point of interest. An advantage of ensembling is its greater flexibility compared to the penetrance modification approach. Ensembling can easily handle any number of models that can be of any form. Including additional risk factors is straightforward. The penetrance modification model requires more assumptions, as it specifically combines a model based on Mendelian inheritance with a relative hazard model. Additional risk modifiers can be incorporated as new relative hazard estimates become available; however, it is important to properly scale the relative hazards to be compatible with the hazard functions they are meant to modify and to consider whether the effects of the risk modifiers differ by carrier status. BRCAPRO + BCRAT (M) currently modifies only the non-carrier hazard function, for which it uses the BCRAT relative hazard. Future work is needed to evaluate the inclusion of modifiers of the carrier hazard functions. One more advantage of ensembling is that after the model is trained it only requires the predictions from the models being combined (along with any additional risk factors that are explicitly included in the ensemble model), and does not require the raw inputs, which are potentially less accessible than the predictions.

BRCAPRO + BCRAT (M) has been externally validated in two average-risk cohorts: a large US mammography screening cohort of over 120,000 women [58] and an Australian cohort of over 7000 women [14]. In the US cohort, BRCAPRO + BCRAT (M) performed better than BRCAPRO, especially among women with a family history of cancer, and per-

formed similarly to BCRAT. In the Australian cohort, BRCAPRO + BCRAT (M) performed similarly to BRCAPRO and worse than BCRAT. To investigate whether these results may have been driven by cohort characteristics, we simulated data to mimic the characteristics of the Australian cohort, as the original data are not publicly available. In these additional simulations, we were unable to reproduce the results reported in the Australian study. Rather than seeing similar performance of BRCAPRO and BRCAPRO + BCRAT (M), as shown in [14], in our simulations BRCAPRO + BCRAT (M) outperformed BRCAPRO alone, which is consistent with the results from the US cohort validation [58] (Appendix E).

While at least two breast cancer risk prediction models integrating detailed family history and non-genetic risk factors, namely, IBIS and BOADICEA, are well-validated and widely available, our work contributes to the improvement of clinical risk prediction in the following ways: (1) we have developed an expanded version of BRCAPRO that can easily be integrated in clinical and research settings where BRCAPRO is currently used, and (2) the model combination methodology proposed in this paper is applicable to a wide range of prediction problems. BRCAPRO is currently part of seven widely used and commercially available clinical tools. These have thousands of users, not all of whom have access to IBIS and BOADICEA. The proposed BRCAPRO + BCRAT (M) model has been added to the software implementation of BRCAPRO, which can facilitate its integration in clinical and research settings where BRCAPRO is used. Two useful features of BRCAPRO are the free availability of its software and code (through the BayesMendel R package) for research purposes and the option for users to specify their own values for the model parameters (through the `brcaparams` function in the BayesMendel R package), including modifications to the allele frequencies and penetrance instead of using the default published estimates. This provides researchers with the flexibility to tailor the model to different populations as well as an easy way to check the sensitivity of risk predictions to differing sets of parameter values. While risk calculation tools for IBIS and BOADICEA are freely available, to the best of our knowledge the code is not easily accessible and there is no option to customize and implement user-specified parameters. Furthermore, the model combination approaches we propose can be applied to many other prediction problems. Ensemble learning can be used to combine an arbitrary number of models of any type. Penetrance modification is a way to incorporate risk factors beyond family history into complex Mendelian models, including the recently developed multi-gene and multi-syndrome model PanelPro [15,62], which can incorporate an arbitrary number of genes and cancer types and can potentially provide greater clinical benefit than single-syndrome models such as BRCAPRO.

## 5. Conclusions

In summary, we have demonstrated the feasibility of two complementary and methodologically different approaches to integrating the popular breast cancer risk assessment models BRCAPRO and BCRAT. Both combination models show accuracy improvements over the existing models, and can potentially prove effective in clinical application. In the CGN validation, BRCAPRO + BCRAT (M) achieved comparable performance to the IBIS model by leveraging the strengths of BRCAPRO and BCRAT. Additional validation of BRCAPRO + BCRAT (M) in independent prospective studies, ideally with larger and more diverse cohorts and more complete covariate information, would provide further support for clinical adoption.

**Author Contributions:** Conceptualization, Z.G., K.H., H.U., L.T., G.P. and D.B.; methodology, Z.G., H.U., L.T., G.P. and D.B.; software, Z.G. and T.H.; formal analysis, Z.G.; resources, A.M.M., K.H., A.S., G.P. and D.B.; data curation, Z.G., T.H. and A.M.M.; writing—original draft preparation, Z.G.; writing—review and editing, Z.G., T.H., A.M.M., K.H., A.S., H.U., L.T., G.P. and D.B.; visualization, Z.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by NSERC PGSD35023622017 (Z.G.); NIH NCI N01PC74400-6-0-1; NIH NCI 5P30CA006516-56 (G.P.); NSF 1810829 (G.P., L.T.); NIH NCI 5T32CA009337-40, NIH T32

CA 009001 (T.H.); Dana-Farber Cancer Institute Research Scientist Development Fund (D.B.); and American Cancer Society Grant 131052-MRSG-17-144-01-CCE (A.M.M.).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** De-identified data that support the findings of this study are available on request. The data are not publicly available due to privacy restrictions. The simulation code is available at [https://github.com/zoeguan/brcapro\\_bcrat\\_combination](https://github.com/zoeguan/brcapro_bcrat_combination) (accessed on 20 July 2020).

**Conflicts of Interest:** Giovanni Parmigiani is a member of the Scientific Advisory Board of Ambry Genetics, a commercial provider of cancer susceptibility tests. Commercial use of the BRCAPRO model is licensed by the Dana Farber Cancer Institute as part of the BayesMendel software package. Licensing revenues are used for model and software upgrades. None of the authors derives personal income from BRCAPRO licensing revenues. Kevin Hughes receives Honoraria from Hologic (Surgical implant for radiation planning with breast conservation and wire free breast biopsy) and Myriad Genetics. He has a financial interest in CRA Health (Formerly Hughes RiskApps), which was recently sold to Volpara. CRA Health develops risk assessment models/software with a particular focus on breast cancer and colorectal cancer. Kevin Hughes is a founder of the company. He is the Co-Creator of Ask2Me.Org which is freely available for clinical use and is licensed for commercial use by the Dana Farber Cancer Institute, Harvard University, and the Massachusetts General Hospital (MGH). His interests in CRA Health and Ask2Me.Org were reviewed and are managed by MGH and Partners Health Care in accordance with their conflict of interest policies.

## Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area under the curve
C	Concordance
CI	Confidence interval
E	Expected
IPCW	Inverse probability of censoring weights
O	Observed
SEER	Surveillance, Epidemiology, and End Results
SNB	Standardized net benefit

## Appendix A. Notation Details

Given a female individual of age  $a$  without a previous diagnosis of breast cancer who presents for risk assessment (the counselee), the goal is to predict her risk of developing invasive breast cancer within  $\tau$  years ( $\tau$  being a pre-specified positive integer) based on family history  $\mathbf{H}$  and other risk factors  $\mathbf{Z}$  (described below) while accounting for death from other causes as a competing risk. Formally, we are interested in the probability  $P(T \leq a + \tau, J = B | T > a, \mathbf{H}, \mathbf{Z})$ , where  $T$  is the age at onset of breast cancer (a non-negative integer) and  $J$  indicates whether she develops breast cancer before dying from other causes ( $J = B$ ) or dies without developing breast cancer ( $J = D$ ). In this paper, we focus on crude or absolute risk of breast cancer, which refers to the risk in the presence of competing causes of death, as opposed to net or pure risk, which refers to the hypothetical risk if competing causes of death were eliminated. BRCAPRO and IBIS provide both the crude and net risks, while BCRAT provides only the crude risk.

While BRCAPRO uses a discrete model for  $T$ , BCRAT and IBIS use continuous models for the time to breast cancer, so we introduce additional notation for these models. Treating time as continuous, let  $\tilde{T}_B$  be the age at onset of breast cancer,  $\tilde{T}_D$  the age at death from other causes, and  $\tilde{T} = \min(\tilde{T}_B, \tilde{T}_D)$ , with  $\tilde{T}_B$ ,  $\tilde{T}_D$ , and  $\tilde{T}$  taking on values in the interval  $[0, \infty)$ .  $T_B = \lfloor \tilde{T}_B \rfloor$  and  $T_D = \lfloor \tilde{T}_D \rfloor$ , and  $\tilde{T} = \lfloor T \rfloor$  (where  $\lfloor \cdot \rfloor$  denotes the floor function).

We use discrete  $T$  for the penetrance modification model, as it is an extension of BRCAPRO. For the ensemble model, time can be treated as either continuous or discrete. Without loss of generality, we use continuous time notation.

Appendix A.1. BRCAPRO

Family history can be represented as a pedigree, or a graph where each node is a family member and edges flow from parents to offspring. The family history information used by BRCAPRO,  $\mathbf{H}$ , includes the following variables for each family member (missing values are allowed): gender, breast cancer status, age of onset of breast cancer if applicable, ovarian cancer status, age of onset of ovarian cancer if applicable, and current age or age at death. Information on preventative interventions (mastectomy/oophorectomy) and genetic testing results for BRCA1/BRCA2 can be included as well.

Let  $n$  be the number of family members in the pedigree and let  $i = 1, \dots, n$  index the family members, where  $i = 1$  corresponds to the counselee. We now describe the notation for the cancer histories of the family members, which is slightly different from the notation for the counselee above because the carrier probability calculation in BRCAPRO incorporates ovarian cancer as well as breast cancer and treats death from other causes as a form of censoring rather than a competing risk. Let  $T_{Bi}$  be the age of breast cancer onset,  $T_{Oi}$  the age of ovarian cancer onset,  $C_i$  the current age (if alive) or death age (if dead),  $X_{Bi} = \min(T_{Bi}, C_i)$ ,  $X_{Oi} = \min(T_{Oi}, C_i)$ ,  $\delta_{Bi} = I(T_{Bi} \leq C_i)$ , and  $\delta_{Oi} = I(T_{Oi} \leq C_i)$ . Denote family member  $i$ 's cancer history by  $\mathbf{H}_i = (X_{Bi}, X_{Oi}, \delta_{Bi}, \delta_{Oi})$ . Let  $\mathbf{G}_i = (G_{i1}, G_{i2})$  be the genotype of family member  $i$ , where  $G_{i1}$  indicates BRCA1 carrier status and  $G_{i2}$  indicates BRCA2 carrier status, and let  $U_i$  indicate whether family member  $i$  is male. Let  $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_n)$  and  $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_n)$ .

Using Bayes' rule and the assumption of conditional independence of phenotypes assigned genotypes (across individuals as well as cancer types), the counselee's probability of having genotype  $\mathbf{G}_1$  is

$$P(\mathbf{G}_1 | \mathbf{H}, \mathbf{U}) = \frac{P(\mathbf{G}_1) \sum_{\mathbf{G}_2, \dots, \mathbf{G}_n} \prod_{i=1}^n P(\mathbf{H}_i | \mathbf{G}_i, U_i) P(\mathbf{G}_2, \dots, \mathbf{G}_n | \mathbf{G}_1)}{\sum_{\mathbf{G}_1} P(\mathbf{G}_1) \sum_{\mathbf{G}_2, \dots, \mathbf{G}_n} \prod_{i=1}^n P(\mathbf{H}_i | \mathbf{G}_i, U_i) P(\mathbf{G}_2, \dots, \mathbf{G}_n | \mathbf{G}_1)}. \tag{A1}$$

The summation over genotypes is calculated using the Elston–Stewart peeling algorithm [63], and  $P(\mathbf{G}_2, \dots, \mathbf{G}_n | \mathbf{G}_1)$  is calculated based on Mendelian laws of inheritance. The prevalences  $P(\mathbf{G}_i)$  ( $P(\mathbf{G}_1) = P(\mathbf{G}_2) = \dots = P(\mathbf{G}_n)$ ) are obtained from the literature and are ethnicity-specific (in particular, different prevalences are used for Ashkenazi Jewish and non-Ashkenazi Jewish families);  $P(\mathbf{H}_i | \mathbf{G}_i, U_i)$  is calculated using literature-based net penetrances for breast and ovarian cancer. The penetrances are functions of age, and are cancer- and sex-specific. The penetrance functions for non-carriers are based on rates from the Surveillance, Epidemiology, and End Results (SEER) program and are race-specific, while the penetrance functions for carriers are from a meta-analysis of published studies [34].

After estimating the carrier probabilities, BRCAPRO calculates future risk of breast and ovarian cancer through a weighted average of the genotype-specific risks. The counselee's crude risk of developing breast cancer between ages  $a$  and  $a + \tau$ , conditional on having genotype  $\mathbf{g}$  ( $\mathbf{g} \in \{(0,0), (1,0), (0,1)\}$ ) and not having breast cancer by age  $a$ , is

$$\begin{aligned} P(T \leq a + \tau, J = B | T > a, \mathbf{G}_1 = \mathbf{g}) &= \sum_{t=a+1}^{a+\tau} \frac{P(T = t, J = B | T \geq t, \mathbf{G}_1 = \mathbf{g}) P(T \geq t | \mathbf{G}_1 = \mathbf{g})}{P(T > a | \mathbf{G}_1 = \mathbf{g})} \\ &= \sum_{t=a+1}^{a+\tau} \lambda_B^{\mathbf{g}}(t) \prod_{u=a+1}^{t-1} (1 - \lambda_B^{\mathbf{g}}(u) - \lambda_D(u)) \end{aligned} \tag{A2}$$

where  $\lambda_B^{\mathbf{g}}(t) = P(T = t, J = B | T \geq t, \mathbf{G}_1 = \mathbf{g})$  is the cause-specific hazard of breast cancer conditional on genotype  $\mathbf{g}$  and  $\lambda_D(t) = P(T = t, J = D | T \geq t)$  is the cause-specific hazard of death from causes other than breast cancer;  $\lambda_B^{\mathbf{g}}(t)$  is calculated from the crude female breast cancer penetrance for genotype  $\mathbf{g}$ ,  $P(T = t, J = B | \mathbf{G}_1 = \mathbf{g})$  using the recursive formula

$$\lambda_B^{\mathbf{g}}(t) = \frac{P(T = t, J = B | \mathbf{G}_1 = \mathbf{g})}{\prod_{u=1}^{t-1} (1 - \lambda_B^{\mathbf{g}}(u) - \lambda_D(u))}, \tag{A3}$$

while  $\lambda_D(t)$  is estimated based on SEER mortality rates for all causes except breast cancer. The final risk estimate is

$$P(T \leq a + \tau, J = B | T > a, \mathbf{H}, \mathbf{U}) = \sum_{\mathbf{g}} P(T \leq a + \tau, J = B | T > a, \mathbf{G}_1 = \mathbf{g}) P(\mathbf{G}_1 = \mathbf{g} | \mathbf{H}, \mathbf{U}). \tag{A4}$$

Appendix A.2. BCRAT

BCRAT [4–7] estimates the relative hazard of developing breast cancer based on age (dichotomized into  $< 50$  and  $\geq 50$ ) and the following risk factors:  $Z_1$  = age at menarche,  $Z_2$  = number of benign breast biopsies,  $Z_3$  = age at first live birth (set  $Z_3 = 25$  for nulliparous non-Hispanic women and  $Z_3 = 30$  for nulliparous Hispanic women),  $Z_4$  = number of female first-degree relatives with breast cancer, and  $Z_5$  = presence of atypical hyperplasia in a breast biopsy (0, 1, or unknown). Let  $\mathbf{Z} = (Z_1, \dots, Z_5)$ .

The relative hazard for an individual of age  $t$  with risk factors  $\mathbf{Z}$  compared to an individual of age  $t$  with no BCRAT risk factors (other than age) is

$$\begin{aligned} r(t, \mathbf{Z}) = & \exp(\beta_1 I[Z_1 \in [12, 13]] + \beta_2 I[Z_1 < 12] + \\ & \beta_3 I[Z_2 = 1] + \beta_4 I[Z_2 \geq 2] + \beta_5 I[t \geq 50] I[Z_2 = 1] + \beta_6 I[t \geq 50] I[Z_2 \geq 1] + \\ & \beta_7 I[Z_3 \in [20, 24]] + \beta_8 I[Z_3 \in [25, 29]] + \beta_9 I[Z_3 > 29] + \\ & \beta_{10} I[Z_4 = 1] + \beta_{11} I[Z_4 = 2] + \\ & \beta_{12} I[Z_3 \in [20, 24]] I[Z_4 = 1] + \beta_{13} I[Z_3 \in [25, 29]] I[Z_4 = 1] + \beta_{14} I[Z_3 > 29] I[Z_4 = 1] + \\ & \beta_{15} I[Z_3 \in [20, 24]] I[Z_4 \geq 2] + \beta_{16} I[Z_3 \in [25, 29]] I[Z_4 \geq 2] + \beta_{17} I[Z_3 > 29] I[Z_4 \geq 2] + \\ & \beta_{18} I[Z_2 > 0] I[Z_5 = 0] + \beta_{19} I[Z_2 > 0] I[Z_5 = 1]), \end{aligned} \tag{A5}$$

where  $I[\cdot]$  denotes the indicator function (equal to 1 if the bracketed expression is true and 0 otherwise). The relative hazard model includes interactions between age and number of biopsies, as well as age at first live birth and number of affected relatives. The regression coefficients were estimated from U.S. case-control studies. Separate models were fit to data from White, African-American, Asian, and Hispanic women to obtain race/ethnicity-specific estimates.

The crude risk of developing breast cancer between ages  $a$  and  $a + \tau$ , conditional on not having breast cancer at age  $a$ , is

$$P(\tilde{T} \leq a + \tau, J = B | \tilde{T} > a, \mathbf{Z}) = \int_a^{a+\tau} \tilde{\lambda}_{B,0}(t) r(t, \mathbf{Z}) \exp\left\{-\int_a^t (\tilde{\lambda}_{B,0}(u) r(u, \mathbf{Z}) + \tilde{\lambda}_D(u)) du\right\} dt, \tag{A6}$$

where  $\tilde{\lambda}_{B,0}(t) = \lim_{dt \rightarrow 0} P(t \leq \tilde{T} < t + dt, J = B | \tilde{T} \geq t, \mathbf{Z} = \mathbf{0}) / dt$  is the cause-specific hazard of breast cancer for those with no BCRAT risk factors and  $\tilde{\lambda}_D(u) = \lim_{dt \rightarrow 0} P(t \leq \tilde{T} < t + dt, J = D | \tilde{T} \geq t) / dt$  is the cause-specific hazard of death from causes other than breast cancer;  $\tilde{\lambda}_{B,0}(t)$  is calculated from  $\tilde{\lambda}_B(t) = \lim_{dt \rightarrow 0} P(t \leq \tilde{T} < t + dt, J = B | \tilde{T} \geq t) / dt$ , the cause-specific hazard of breast cancer in the general population, using the formula

$$\tilde{\lambda}_{B,0}(t) = \tilde{\lambda}_B(t)(1 - AR(t)), \tag{A7}$$

where

$$AR(t) = 1 - \frac{1}{\sum_{\mathbf{z}} r(t, \mathbf{z}) P(\mathbf{Z} = \mathbf{z} | t)} \tag{A8}$$

is the population attributable risk due to  $\mathbf{Z}$  for those of age  $t$  [64] and  $P(\mathbf{Z} = \mathbf{z} | t)$  is the probability mass function for  $\mathbf{Z}$  among individuals of age  $t$  in the general population. Here,  $P(\mathbf{Z} = \mathbf{z} | t)$  and  $r(t, \mathbf{z})$  are both assumed to be constant for  $t < 50$  and for  $t \geq 50$ , meaning that  $AR(t)$  is as well. Race-specific estimates of  $\tilde{\lambda}_B(t)$  and  $AR(t)$  were obtained from SEER

data, and  $\tilde{\lambda}_D(t)$  was estimated based on SEER mortality rates for all causes except breast cancer. In the implementation of the model, the age scale is divided into thirteen intervals, with  $\tilde{\lambda}_B(t)$  and  $\tilde{\lambda}_D(t)$  assumed to be constant on each interval; see [4] for more details.

Appendix A.3. IBIS

IBIS calculates a weighted average of the cumulative penetrances for each genotype:

$$P(\tilde{T} \leq t, J = B | \mathbf{H}, \mathbf{U}) = \sum_{\mathbf{g}, \theta} P(\tilde{T} \leq t, J = B | \mathbf{G}, \Theta) P(\mathbf{G}_1 = \mathbf{g}, \Theta = \theta | \mathbf{H}, \mathbf{U}), \tag{A9}$$

where  $\Theta$  indicates the counselee’s carrier status with respect to the hypothetical gene. In IBIS, carriers of both BRCA1 and BRCA2 mutations are treated as BRCA1 carriers.

Let  $\mathbf{V}$  represent the following risk factors: age at menarche, age at menopause, height, body mass index, age at first live birth, menopausal hormone therapy, atypical hyperplasia, lobular carcinoma in situ, breast density, and polygenic risk score. The crude risk of developing breast cancer between ages  $a$  and  $a + \tau$ , conditional on not having breast cancer at age  $a$ , is

$$P(\tilde{T} \leq a + \tau, J = B | \tilde{T} > a, \mathbf{H}, \mathbf{U}, \mathbf{V}) = \int_a^{a+\tau} \tilde{\lambda}_{B,\mathbf{H}}(t) r_P(\mathbf{V}) \times \exp\left\{-\int_a^t (\tilde{\lambda}_{B,\mathbf{H}}(u) r_P(\mathbf{V}) + \tilde{\lambda}_D(u)) du\right\} dt, \tag{A10}$$

where  $\tilde{\lambda}_{B,\mathbf{H}}(t) = \lim_{dt \rightarrow 0} P(t \leq \tilde{T} < t + dt, J = B | \tilde{T} \geq t, \mathbf{H}, \mathbf{U}) / dt$  is the cause-specific hazard of breast cancer conditional on family history, and  $r_P(\mathbf{V})$  is a normalized version of the relative hazard of breast cancer associated with risk factors  $\mathbf{V}$

$$r_P(\mathbf{V}) = \phi(\mathbf{V})(1 - AR) = \frac{\phi(\mathbf{V})}{\int \phi(\mathbf{v}) f(\mathbf{v}) d\mathbf{v}} \tag{A11}$$

where  $\phi(\mathbf{V})$  is the relative hazard associated with  $\mathbf{V}$  (relative to the no-risk population),  $AR$  is the population attributable risk due to  $\mathbf{V}$ , and  $f(\mathbf{v})$  is the population prevalence of  $\mathbf{v}$ . In the IBIS model, the risk factors are assumed to be independent and  $r_P(\mathbf{V})$  is approximated by

$$r_P(\mathbf{V}) \approx \prod_k \frac{\phi(V_k)}{\int \phi(v_k) f(v_k) dv_k} \tag{A12}$$

where  $k$  indexes the risk factors in  $\mathbf{V}$ .

Appendix A.4. Penetrance Modification Model: BRCAPRO + BCRAT (M)

The penetrance modification approach combines BRCAPRO and BCRAT by using the BCRAT relative hazard to modify the genotype-specific hazard functions in BRCAPRO. Because BCRAT tends to underestimate risk for known carriers of BRCA1/2 mutations and is not recommended for these individuals [37], we propose to apply the BCRAT relative hazard only to the non-carrier hazard function in BRCAPRO.

Analogous to the BCRAT relative hazard model

$$\tilde{\lambda}_B(t | \mathbf{Z}) = \tilde{\lambda}_{B,0}(t) r(t, \mathbf{Z}), \tag{A13}$$

where  $\tilde{\lambda}_B(t | \mathbf{Z}) = \lim_{dt \rightarrow 0} P(t \leq \tilde{T} < t + dt, J = B | \tilde{T} \geq t, \mathbf{Z}) / dt$  is the cause-specific hazard of breast cancer conditional on  $\mathbf{Z}$ , we can define the following relative hazard model for non-carriers as

$$\tilde{\lambda}_B(t | \mathbf{Z}, \mathbf{G}_1 = \mathbf{0}) = \tilde{\lambda}_B^0(t) r^0(t, \mathbf{Z}), \tag{A14}$$

where  $\tilde{\lambda}_B(t|\mathbf{Z}, \mathbf{G}_1 = \mathbf{0}) = \lim_{dt \rightarrow 0} P(t \leq \tilde{T} < t + dt, J = B | \tilde{T} \geq t, \mathbf{Z}, \mathbf{G}_1 = \mathbf{0})/dt$  is the non-carrier cause-specific hazard of breast cancer conditional on  $\mathbf{Z}$ ,  $\tilde{\lambda}_B^0(t) = \lim_{dt \rightarrow 0} P(t \leq \tilde{T} < t + dt, J = B | \tilde{T} \geq t, \mathbf{G}_1 = \mathbf{0})/dt$  is the non-carrier cause-specific hazard of breast cancer, and  $r^0(t, \mathbf{Z})$  is the relative hazard of breast cancer compared to non-carriers (discussed in more detail below). Models (A13) and (A14) are continuous-time models. To apply the hazard modification in the discrete-time framework of BRCAPRO, we consider the discrete-time analogue induced by Equation (A14) under the setting where only integer-valued  $t$  is observed (see Chapter 2.4.2 of [65]):

$$\lambda_B(t|\mathbf{Z}, \mathbf{G}_1 = \mathbf{0}) = 1 - (1 - \lambda_B^0(t))^{r^0(t, \mathbf{Z})}, \tag{A15}$$

where  $\lambda_B(t|\mathbf{Z}, \mathbf{G}_1 = \mathbf{0}) = P(T = t, J = B | T \geq t, \mathbf{Z}, \mathbf{G}_1 = \mathbf{0})$  is the discrete non-carrier cause-specific hazard of breast cancer conditional on  $\mathbf{Z}$  and  $\lambda_B^0(t)$  is the discrete non-carrier cause-specific hazard of breast cancer defined in Section 2.2.1.

We then modify the calculation of the non-carrier risk in BRCAPRO by replacing  $\lambda_B^0(t)$  in Equation (A2) with  $\lambda_B(t|\mathbf{Z}, \mathbf{G}_1 = \mathbf{0})$  to obtain

$$P(T \leq a + \tau, J = B | T > a, \mathbf{G}_1 = \mathbf{0}, \mathbf{Z}) = \sum_{t=a+1}^{a+\tau} \left( 1 - (1 - \lambda_B^0(t))^{r^0(t, \mathbf{Z})} \right) \times \prod_{u=a+1}^{t-1} \left( (1 - \lambda_B^0(u))^{r^0(u, \mathbf{Z})} - \lambda_D(u) \right). \tag{A16}$$

As in BRCAPRO, the final risk is a weighted average of the genotype-specific risks:

$$P(T \leq a + \tau, J = B | T > a, \mathbf{H}, \mathbf{U}, \mathbf{Z}) = \sum_{\mathbf{g}} P(T \leq a + \tau, J = B | T > a, \mathbf{G}_1 = \mathbf{g}, \mathbf{Z}) P(\mathbf{G}_1 = \mathbf{g} | \mathbf{H}, \mathbf{U}). \tag{A17}$$

Similar to  $r_P(T)$  from IBIS (Equation (A11)),  $r^0(t, \mathbf{Z})$  is a normalized version of  $r(t, \mathbf{Z})$  where the normalization factor is the average relative hazard among non-carriers:

$$r^0(t, \mathbf{Z}) = r(t, \mathbf{Z})(1 - AR^0(t)) = \frac{r(t, \mathbf{Z})}{\sum_{\mathbf{z}} r(t, \mathbf{z}) P(\mathbf{Z} = \mathbf{z} | t, \mathbf{G}_1 = \mathbf{0})} \tag{A18}$$

where  $AR^0(t)$  is the population-attributable risk due to  $\mathbf{Z}$  among non-carriers of age  $t$ . Normalization is necessary because  $r(t, \mathbf{Z})$  is with respect to the no-risk population; in order to modify  $\lambda_B^0(t)$ , we need to obtain the relative hazard with respect to the non-carrier population.

Because the prevalence of BRCA1/2 mutations in the general population is very low (estimated at 1/400 [66]), we approximate  $AR^0(t)$  with  $AR(t)$ , i.e., we assume  $P(\mathbf{Z} = \mathbf{z} | t, \mathbf{G}_1 = \mathbf{0}) \approx P(\mathbf{Z} = \mathbf{z} | t)$ . Therefore, BRCAPRO + BCRAT (M) takes parameters from existing models and does not need to be trained on new data (although the parameters should be updated as new data becomes available). Race-specific estimates of  $AR(t)$  are available from BCRAT; however, as they are based on data from the 1980s to early 2000s, we re-estimated  $AR(t)$  based on the distribution of BCRAT covariates in more recent data from the 2015 National Health Interview Survey (NHIS), which uses a cross-sectional sample of U.S. adults designed to be representative of the U.S. general population. As in IBIS (Equation (A12)) we assumed that the risk factors are independent, except that we used the joint distribution of age at first live birth and number of affected first-degree relatives, which was because Equation (A5) includes an interaction between these variables. The race-specific estimates from the NHIS are provided in Appendix B.

Because the BRCAPRO carrier probabilities  $P(\mathbf{G}_1 = \mathbf{g}|\mathbf{H})$  already account for family history, it may seem redundant to include the BCRAT family history variable (number of female first-degree relatives with breast cancer) among the penetrance-modifying risk factors (Equation (A17)). However, its inclusion can be useful because (1) there is a strong interaction between the family history variable and age at first live birth in BCRAT and (2) the BCRAT family history variable could potentially account for residual familial risk due to shared environmental factors or genes other than BRCA1/BRCA2, which are not currently considered by BRCAPRO.

#### Appendix A.5. Ensemble Model (BRCAPRO + BCRAT (E))

Let  $F_B(\tau) = P(\tilde{T} \leq a + \tau, J = B | \mathbf{T} > a, \mathbf{H}, \mathbf{U}, \mathbf{Z})$ , and let  $F_{B,1}(\tau)$  be the  $\tau$ -year BRCAPRO risk prediction and  $F_{B,2}(\tau)$  the  $\tau$ -year BCRAT risk prediction. For BRCAPRO + BCRAT (E), we combine the  $\tau$ -year BRCAPRO and BCRAT predictions using the logistic regression model

$$\log \frac{F_B(\tau)}{1 - F_B(\tau)} = \beta_0 + \beta_1 F_{B,1}(\tau) + \beta_2 F_{B,2}(\tau) + \beta_3 F_{B,1}(\tau) F_{B,2}(\tau). \quad (\text{A19})$$

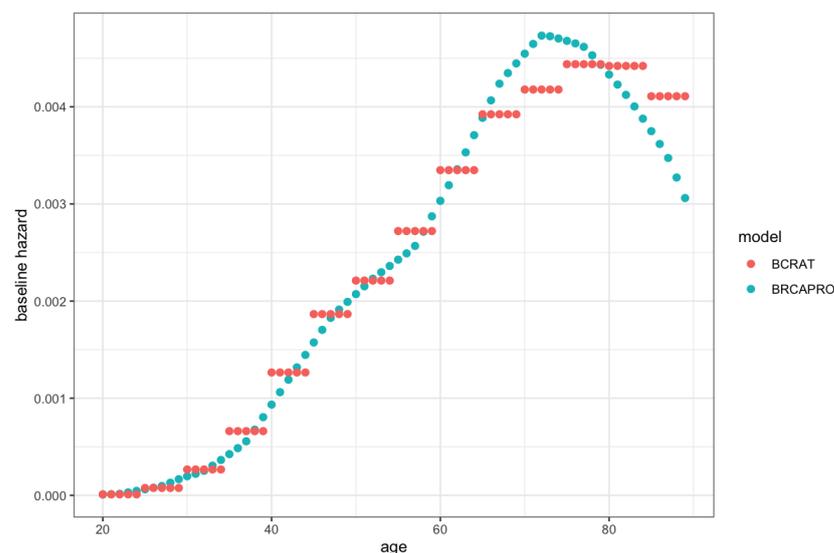
For BRCAPRO + BCRAT (E2), we consider a Fine–Gray proportional subdistribution hazards model [67] for breast cancer, accounting for death as a competing risk:

$$1 - F_B(\tau) = (1 - F_{B,0}(\tau))^{\exp(\beta_1 F_{B,1}(\tau^*) + \beta_2 F_{B,2}(\tau^*) + \beta_3 F_{B,1}(\tau^*) F_{B,2}(\tau^*))} \quad (\text{A20})$$

where  $F_{B,0}(t)$  is the baseline cumulative incidence function for breast cancer and  $\tau^*$  is a predefined time point; the covariates  $F_{B,1}(\tau^*)$  and  $F_{B,2}(\tau^*)$  are calculated at baseline do not vary with  $\tau$ .

## Appendix B. Penetrance Modification Parameters

### Appendix B.1. Hazard Functions in BRCAPRO and BCRAT



**Figure A1.** BCRAT cause-specific hazard of breast cancer for White women in the general population ( $\tilde{\lambda}_B(t) = \tilde{\lambda}_{B,0}(t)/(1 - AR(t))$ ) and BRCAPRO cause-specific hazard of breast cancer for White female non-carriers ( $\lambda_B^0(t)$ ).

Appendix B.2. Population Attributable Risk Estimates for BCRAT Covariates

Table A1. Estimates of  $1/(1 - AR(t))$  from NHIS 2015.

	White	Black	Hispanic	Asian	Native American
< 50	1.81	1.41	1.37	2.10	1.55
≥ 50	1.96	1.44	1.41	2.43	1.94

Appendix C. Ensemble Weights from NWH

For BRCAPRO + BCRAT (E), we fit the logistic regression model

$$\log \frac{F_B(\tau)}{1 - F_B(\tau)} = \beta_0 + \beta_1 \sqrt{F_{B,1}(\tau)} + \beta_2 \sqrt{F_{B,2}(\tau)} + \beta_3 \sqrt{F_{B,1}(\tau)} \sqrt{F_{B,2}(\tau)},$$

to the NWH cohort, where  $\tau = 5$ . The estimated weights are provided below.

Table A2. Coefficient estimates for BRCAPRO + BCRAT (E).

	Estimate	Standard Error
$\hat{\beta}_0$	2.55	1.31
$\hat{\beta}_1$	0.86	0.32
$\hat{\beta}_2$	1.21	0.28
$\hat{\beta}_3$	0.11	0.06

For BRCAPRO + BCRAT (E2), we fit the Fine–Gray model

$$1 - F_B(\tau) = (1 - F_{B,0}(\tau)) \exp(\beta_1 \sqrt{F_{B,1}(\tau^*)} + \beta_2 \sqrt{F_{B,2}(\tau^*)} + \beta_3 \sqrt{F_{B,1}(\tau^*)} \sqrt{F_{B,2}(\tau^*)})$$

to the NWH cohort with  $\tau^* = 5$ . The estimated weights are provided below.

Table A3. Coefficient estimates for BRCAPRO + BCRAT (E2).

	Estimate	Standard Error
$\hat{\beta}_1$	19.76	5.67
$\hat{\beta}_2$	24.18	6.22
$\hat{\beta}_3$	−113.06	45.30

Appendix D. CGN Validation: Additional Tables

Table A4. CGN cohort characteristics by center.

Variable Category	N	Age (Median [IQR])	Affected 1st-Degree Relatives (%)			Follow-Up (Median [IQR])	Censored (%)	5-Year Cases (%)
			0	1	2+			
BAYLOR	69	47 [38, 52]	34 (49.3)	28 (40.6)	7 (10.1)	7.5 [6.1, 8.6]	12 (17.4)	0 (0.0)
CU	1198	51 [41, 64]	528 (44.1)	547 (45.7)	123 (10.3)	7.6 [6.4, 8.5]	85 (7.1)	23 (1.9)
DUKE	286	46 [38, 53]	134 (46.9)	116 (40.6)	36 (12.6)	7.2 [6.2, 8.2]	40 (14.0)	9 (3.1)
EMORY	136	44 [38.8, 51]	56 (41.2)	51 (37.5)	29 (21.3)	7.2 [6.6, 8.3]	29 (21.3)	3 (2.2)
GU	309	43 [35, 52]	107 (34.6)	147 (47.6)	55 (17.8)	7.8 [6.3, 8.5]	81 (26.2)	2 (0.6)
JH	469	47 [39, 56]	279 (59.5)	148 (31.6)	42 (9.0)	8.0 [6.1, 9.0]	73 (15.6)	10 (2.1)
MDAND	295	45 [37, 53]	215 (72.9)	64 (21.7)	16 (5.4)	6.1 [4.2, 7.0]	87 (29.5)	4 (1.4)
UCI	608	48 [37, 59]	352 (57.9)	223 (36.7)	33 (5.4)	5.3 [4.0, 7.1]	209 (34.4)	4 (0.7)
UNC	229	46 [39, 53]	80 (34.9)	111 (48.5)	38 (16.6)	8.0 [7.1, 9.0]	28 (12.2)	6 (2.6)
UNM	324	[41, 63]	160 (49.4)	123 (38.0)	41 (12.7)	6.6 [6.0, 7.5]	43 (13.3)	11 (3.4)
UPENN	540	45 [37, 53]	297 (55.0)	185 (34.3)	58 (10.7)	8.1 [6.6, 9.1]	76 (14.1)	8 (1.5)
UTAH	880	47 [35, 61]	522 (59.3)	298 (33.9)	60 (6.8)	7.3 [5.4, 8.0]	61 (6.9)	14 (1.6)
UTSA	92	43 [35.8, 52]	29 (31.5)	48 (52.2)	15 (16.3)	5.1 [4.0, 6.5]	36 (39.1)	1 (1.1)
UTSW	247	41 [33.5, 47]	88 (35.6)	116 (47.0)	43 (17.4)	6.6 [5.6, 7.6]	30 (12.1)	4 (1.6)
UWASH	1632	46 [37, 56]	1290 (79.0)	291 (17.8)	51 (3.1)	7.6 [6.9, 8.3]	44 (2.7)	13 (0.8)

**Table A5.** Performance over follow-up period in entire CGN cohort. B+B: BRCAPRO + BCRAT. ΔBS: % relative improvement in Brier Score compared to BRCAPRO. ΔLS: % relative improvement in logarithmic score compared to BRCAPRO. The “Comparisons Across Bootstrap Replicates” section shows pairwise comparisons involving the combination models across 1000 bootstrap replicates of the validation dataset; the row for  $A > B$  shows the proportion of bootstrap replicates where model A outperformed model B with respect to each metric. Proportions  $> 0.5$  are highlighted in blue (with darker shades of blue for higher proportions) and proportions  $\leq 0.5$  are highlighted in red (with darker shades of red for lower proportions).

	O/E	C	ΔBS	ΔLS
<b>Performance Metrics</b>				
B+B (M)	1.04 (0.88, 1.19)	0.66 (0.61, 0.70)	0.15 (−0.61, 0.84)	0.49 (−0.04, 1.04)
B+B (E2)	1.16 (0.98, 1.33)	0.65 (0.60, 0.70)	0.42 (−0.05, 0.89)	
BRCAPRO	1.33 (1.12, 1.52)	0.61 (0.56, 0.66)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
BCRAT	1.14 (0.96, 1.30)	0.65 (0.61, 0.69)	0.28 (−0.55, 1.09)	−0.74 (−1.69, 0.11)
IBIS	0.99 (0.83, 1.13)	0.64 (0.59, 0.69)	−0.30 (−1.00, 0.33)	
<b>Comparisons Across Bootstrap Replicates</b>				
B+B(M)>B+B(E2)	0.892	0.788	0.105	
B+B(M)>BRCAPRO	0.975	0.997	0.668	0.956
B+B(M)>BCRAT	0.866	0.721	0.279	1.000
B+B(M)>IBIS	0.445	0.775	0.928	
B+B(E2)>BRCAPRO	0.993	0.999	0.958	
B+B(E2)>BCRAT	0.044	0.504	0.724	
B+B(E2)>IBIS	0.172	0.663	0.988	

**Table A6.** Performance over follow-up period in CGN cohort stratified by family history (whether or not the proband met the NCCN criteria for further genetic risk evaluation [55]); in applying the criteria, we only used information on breast and ovarian cancer diagnoses in relatives. B+B: BRCAPRO + BCRAT. ΔBS: % relative improvement in Brier Score compared to BRCAPRO. The “Comparisons Across Bootstrap Replicates” section shows pairwise comparisons involving the combination models across 1000 bootstrap replicates of the validation dataset; the row for  $A > B$  shows the proportion of bootstrap replicates where model A outperformed model B with respect to each metric. Proportions  $> 0.5$  are highlighted in blue (with darker shades of blue for higher proportions) and proportions  $\leq 0.5$  are highlighted in red (with darker shades of red for lower proportions).

	O/E	C	ΔBS	ΔLS
<b>Strong Family History (45 cases)</b>				
<b>Performance Metrics</b>				
B+B (M)	0.76 (0.54, 1.01)	0.65 (0.52, 0.74)	−0.67 (−3.29, 1.04)	−0.08 (−2.77, 2.26)
B+B (E2)	1.03 (0.72, 1.35)	0.62 (0.48, 0.71)	0.15 (−1.17, 1.32)	
BRCAPRO	1.27 (0.89, 1.68)	0.62 (0.49, 0.72)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
BCRAT	0.95 (0.66, 1.26)	0.57 (0.43, 0.66)	−0.50 (−3.19, 1.72)	−3.45 (−7.02, 0.03)
IBIS	0.70 (0.49, 0.92)	0.66 (0.51, 0.76)	−1.49 (−4.19, 0.48)	
<b>Comparisons Across Bootstrap Replicates</b>				
B+B(M)>B+B(E2)	0.233	0.969	0.101	
B+B(M)>BRCAPRO	0.550	0.855	0.223	0.461
B+B(M)>BCRAT	0.146	0.976	0.393	0.999
B+B(M)>IBIS	0.987	0.327	0.812	
B+B(E2)>BRCAPRO	0.825	0.487	0.559	
B+B(E2)>BCRAT	0.527	0.964	0.839	
B+B(E2)>IBIS	0.855	0.094	0.931	
B+B (M)	1.21 (0.99, 1.42)	0.66 (0.61, 0.71)	0.47 (−0.08, 1.11)	0.61 (0.16, 1.06)
B+B (E2)	1.23 (1.01, 1.44)	0.65 (0.59, 0.70)	0.52 (0.08, 1.00)	
BRCAPRO	1.35 (1.11, 1.58)	0.60 (0.53, 0.66)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
<b>Less Family History (114 cases)</b>				
<b>Performance Metrics</b>				
BCRAT	1.24 (1.02, 1.45)	0.67 (0.62, 0.72)	0.59 (0.03, 1.23)	−0.17 (−1.01, 0.62)
IBIS	1.18 (0.97, 1.38)	0.65 (0.58, 0.70)	0.16 (−0.35, 0.65)	
B+B(M)>B+B(E2)	0.971	0.768	0.358	
B+B(M)>BRCAPRO	0.995	0.987	0.946	0.996
B+B(M)>BCRAT	0.976	0.125	0.251	0.985
<b>Comparisons Across Bootstrap Replicates</b>				
B+B(M)>IBIS	0.043	0.625	0.832	
B+B(E2)>BRCAPRO	0.995	1.000	0.995	
B+B(E2)>BCRAT	0.920	0.087	0.257	
B+B(E2)>IBIS	0.035	0.522	0.934	

## Appendix E. Simulations Based on Australian Validation Study

Using simulations to mimic an Australian validation study [68], we compared the performance of IBIS, BRCAPRO, BCRAT, and BRCAPRO + BCRAT in predicting 15-year risk of breast cancer. We sampled baseline covariates from a real dataset of 40,000 women who underwent mammography screening at Newton-Wellesley Hospital (NWH) from 2007–2009 (described in [16]). The sampled subset consisted of 7608 women with a similar risk factor distribution as the original Australian cohort. We ran IBIS, BRCAPRO, BCRAT, BRCAPRO + BCRAT to predict 15-year risk of breast cancer for these women. Because information on number of biopsies and atypical hyperplasia was not available in the study of Li et al., these variables were set to unknown when running BCRAT and BRCAPRO + BCRAT. We then simulated 15-year breast cancer outcomes based on the 15-year IBIS predictions and compared the performance of the models. Cohort characteristics are shown in Table A7, the performance results from the simulations are shown in Table A8, and the performance results with the original Australian cohort are shown in Table A9.

**Table A7.** Characteristics of simulated and original datasets. Family history is defined as family history of breast cancer in first- or second-degree relatives.

Variable	Simulated	Original
N	7608	7608
Mean age	54.1	58.5
Mean height (cm)	163	162
Mean weight (kg)	70.6	71.3
Mean menarche age	12.8	12.9
Mean age at first birth	26.1	25.4
Mean age at menopause	48.7	49.5
15-year cases (%)	412 (5.4)	351 (4.6)
% with family history	24.9	22.6

**Table A8.** Performance of IBIS, BRCAPRO, BCRAT, and BRCAPRO + BCRAT on simulated data.

Model	Expected	Expected/Observed (95% CI)	AUC (95% CI)
IBIS	422.9	0.97 (0.88, 1.06)	0.63 (0.60, 0.66)
BRCAPRO	374.7	1.10 (1.00, 1.20)	0.55 (0.53, 0.58)
BCRAT	355.2	1.16 (1.05, 1.26)	0.57 (0.54, 0.60)
BRCAPRO + BCRAT (M)	382.8	1.08 (0.98, 1.17)	0.58 (0.55, 0.61)

**Table A9.** Performance of IBIS, BRCAPRO, BCRAT, and BRCAPRO + BCRAT in original study.

Model	Expected	Expected/Observed (95% CI)	AUC (95% CI)
IBIS	341.5	0.97 (0.88, 1.08)	0.57 (0.54, 0.61)
BRCAPRO	389.3	1.11 (1.00, 1.23)	0.51 (0.48, 0.54)
BCRAT	327.9	0.93 (0.84, 1.04)	0.54 (0.51, 0.57)
BRCAPRO + BCRAT (M)	389.7	1.11 (1.00, 1.23)	0.51 (0.48, 0.54)

## References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [[CrossRef](#)] [[PubMed](#)]
2. American Cancer Society. Facts and Figures 2020. 2020. Available online: <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2020.html> (accessed on 3 May 2020).
3. Cintolo-Gonzalez, J.A.; Braun, D.; Blackford, A.L.; Mazzola, E.; Acar, A.; Plichta, J.K.; Griffin, M.; Hughes, K.S. Breast cancer risk models: A comprehensive overview of existing models, validation, and clinical applications. *Breast Cancer Res. Treat.* **2017**, *164*, 263–284. [[CrossRef](#)] [[PubMed](#)]
4. Gail, M.H.; Brinton, L.A.; Byar, D.P.; Corle, D.K.; Green, S.B.; Schairer, C.; Mulvihill, J.J. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J. Natl. Cancer Inst.* **1989**, *81*, 1879–1886. [[CrossRef](#)] [[PubMed](#)]
5. Gail, M.H.; Costantino, J.P.; Pee, D.; Bondy, M.; Newman, L.; Selvan, M.; Anderson, G.L.; Malone, K.E.; Marchbanks, P.A.; McCaskill-Stevens, W.; et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J. Natl. Cancer Inst.* **2007**, *99*, 1782–1792. [[CrossRef](#)] [[PubMed](#)]
6. Matsuno, R.K.; Costantino, J.P.; Ziegler, R.G.; Anderson, G.L.; Li, H.; Pee, D.; Gail, M.H. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. *J. Natl. Cancer Inst.* **2011**, *103*, 951–961. [[CrossRef](#)] [[PubMed](#)]

7. Banegas, M.P.; John, E.M.; Slattery, M.L.; Gomez, S.L.; Yu, M.; LaCroix, A.Z.; Pee, D.; Chlebowski, R.T.; Hines, L.M.; Thompson, C.A.; et al. Projecting individualized absolute invasive breast cancer risk in US Hispanic women. *J. Natl. Cancer Inst.* **2016**, *109*, djw215. [[CrossRef](#)]
8. Parmigiani, G.; Berry, D.; Aguilar, O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *Am. J. Hum. Genet.* **1998**, *62*, 145–158. [[CrossRef](#)]
9. Antoniou, A.C.; Pharoah, P.; Smith, P.; Easton, D.F. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br. J. Cancer* **2004**, *91*, 1580. [[CrossRef](#)]
10. Antoniou, A.C.; Cunningham, A.; Peto, J.; Evans, D.; Lalloo, F.; Narod, S.; Risch, H.; Eyfjord, J.; Hopper, J.; Southey, M.; et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br. J. Cancer* **2008**, *98*, 1457–1466. [[CrossRef](#)]
11. Lee, A.; Mavaddat, N.; Wilcox, A.; Cunningham, A.; Carver, T.; Hartley, S.; Babb, d.V.C.; Izquierdo, A.; Simard, J.; Schmidt, M.; et al. BOADICEA: A comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2019**, *21*, 1708–1718.
12. Tyrer, J.; Duffy, S.W.; Cuzick, J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat. Med.* **2004**, *23*, 1111–1130. [[CrossRef](#)] [[PubMed](#)]
13. Terry, M.B.; Liao, Y.; Whittemore, A.S.; Leoce, N.; Buchsbaum, R.; Zeinomar, N.; Dite, G.S.; Chung, W.K.; Knight, J.A.; Southey, M.C.; et al. 10-year performance of four models of breast cancer risk: A validation study. *Lancet Oncol.* **2019**, *20*, 504–517. [[CrossRef](#)]
14. Li, S.X.; Milne, R.L.; Nguyen-Dumont, T.; English, D.R.; Giles, G.G.; Southey, M.C.; Antoniou, A.C.; Lee, A.; Winship, I.; Hopper, J.L.; et al. Prospective Evaluation over 15 Years of Six Breast Cancer Risk Models. *Cancers* **2021**, *13*, 5194. [[CrossRef](#)] [[PubMed](#)]
15. Lee, G.; Liang, J.W.; Zhang, Q.; Huang, T.; Choirat, C.; Parmigiani, G.; Braun, D. Multi-syndrome, multi-gene risk modeling for individuals with a family history of cancer with the novel R package PanelPRO. *Elife* **2021**, *10*, e68699. [[CrossRef](#)] [[PubMed](#)]
16. McCarthy, A.M.; Guan, Z.; Welch, M.; Griffin, M.E.; Sippo, D.A.; Deng, Z.; Coopey, S.B.; Acar, A.; Semine, A.; Parmigiani, G.; et al. Performance of breast cancer risk assessment models in a large mammography cohort. *JNCI J. Natl. Cancer Inst.* **2019**, *112*, 489–497. [[CrossRef](#)]
17. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [[CrossRef](#)]
18. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
19. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
20. Freund, Y.; Schapire, R.E. Experiments with a New Boosting Algorithm. *ICML* **1996**, *96*, 148–156.
21. Dietterich, T.G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.* **2000**, *40*, 139–157. [[CrossRef](#)]
22. Opitz, D.; Maclin, R. Popular ensemble methods: An empirical study. *J. Artif. Intell. Res.* **1999**, *11*, 169–198. [[CrossRef](#)]
23. Kleinberg, E. Stochastic discrimination. *Ann. Math. Artif. Intell.* **1990**, *1*, 207–239. [[CrossRef](#)]
24. Perrone, M.P.; Cooper, L.N. *When Networks Disagree: Ensemble Methods for Hybrid Neural Networks*; Technical Report; Brown University Institution for Brain and Neural Systems: Providence, RI, USA, 1992.
25. Bartlett, P.; Freund, Y.; Lee, W.S.; Schapire, R.E. Boosting the margin: A new explanation for the effectiveness of voting methods. *Ann. Stat.* **1998**, *26*, 1651–1686. [[CrossRef](#)]
26. Kuncheva, L.I. A theoretical study on six classifier fusion strategies. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 281–286. [[CrossRef](#)]
27. Van der Laan, M.J.; Polley, E.C.; Hubbard, A.E. Super learner. *Stat. Appl. Genet. Mol. Biol.* **2007**, *6*. [[CrossRef](#)]
28. Krogh, A.; Vedelsby, J. Neural network ensembles, cross validation, and active learning. In *Advances in Neural Information Processing Systems 7*; MIT Press: Cambridge, MA, USA, 1995; pp. 231–238.
29. Cunningham, P.; Carney, J. Diversity versus quality in classification ensembles based on feature selection. In *Proceedings of the Machine Learning: ECML 2000: 11th European Conference on Machine Learning, Barcelona, Spain, 31 May–2 June 2000*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 109–116.
30. Debray, T.P.; Koffijberg, H.; Nieboer, D.; Vergouwe, Y.; Steyerberg, E.W.; Moons, K.G. Meta-analysis and aggregation of multiple published prediction models. *Stat. Med.* **2014**, *33*, 2341–2362. [[CrossRef](#)]
31. Steyerberg, E.W.; Borsboom, G.J.; van Houwelingen, H.C.; Eijkemans, M.J.; Habbema, J.D.F. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat. Med.* **2004**, *23*, 2567–2586. [[CrossRef](#)]
32. Janssen, K.; Moons, K.; Kalkman, C.; Grobbee, D.; Vergouwe, Y. Updating methods improved the performance of a clinical prediction model in new patients. *J. Clin. Epidemiol.* **2008**, *61*, 76–86. [[CrossRef](#)]
33. Ming, C.; Viassolo, V.; Probst-Hensch, N.; Chappuis, P.O.; Dinov, I.D.; Katapodi, M.C. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. *Breast Cancer Res.* **2019**, *21*, 75. [[CrossRef](#)]
34. Chen, J.; Bae, E.; Zhang, L.; Hughes, K.; Parmigiani, G.; Braun, D.; Rebbeck, T.R. Penetrance of Breast and Ovarian Cancer in Women Who Carry a BRCA1/2 Mutation and Do not Use Risk-Reducing Salpingo-Oophorectomy: An Updated Meta-analysis. *JNCI Cancer Spectr.* **2020**, *4*, pkaa029. [[CrossRef](#)]

35. Chen, S.; Wang, W.; Broman, K.W.; Katki, H.A.; Parmigiani, G. BayesMendel: an R environment for Mendelian risk prediction. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, 1–19. [[CrossRef](#)] [[PubMed](#)]
36. Brentnall, A.R.; Cuzick, J. Risk models for breast cancer and their validation. *arXiv* **2019**, arXiv:1907.02829.
37. Ready, K.; Arun, B. Clinical assessment of breast cancer risk based on family history. *J. Natl. Compr. Cancer Netw.* **2010**, *8*, 1148–1155. [[CrossRef](#)]
38. Milne, R.L.; Antoniou, A.C. Modifiers of breast and ovarian cancer risks for BRCA1 and BRCA2 mutation carriers. *Endocr.-Relat. Cancer* **2016**, *23*, T69–T84. [[CrossRef](#)]
39. Su, T.L.; Jaki, T.; Hickey, G.L.; Buchan, I.; Sperrin, M. A review of statistical updating methods for clinical prediction models. *Stat. Methods Med. Res.* **2018**, *27*, 185–197. [[CrossRef](#)]
40. Sugiyama, M.; Krauledat, M.; MÄzller, K.R. Covariate shift adaptation by importance weighted cross validation. *J. Mach. Learn. Res.* **2007**, *8*, 985–1005.
41. Huang, J.; Gretton, A.; Borgwardt, K.; Schölkopf, B.; Smola, A.J. Correcting sample selection bias by unlabeled data. In *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*; MIT Press: Cambridge, MA, USA, 2007; pp. 601–608.
42. Sugiyama, M.; Suzuki, T.; Nakajima, S.; Kashima, H.; von Büna, P.; Kawanabe, M. Direct importance estimation for covariate shift adaptation. *Ann. Inst. Stat. Math.* **2008**, *60*, 699–746. [[CrossRef](#)]
43. Kanamori, T.; Hido, S.; Sugiyama, M. A least-squares approach to direct importance estimation. *J. Mach. Learn. Res.* **2009**, *10*, 1391–1445.
44. Steyerberg, E.W.; Vickers, A.J.; Cook, N.R.; Gerdts, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology* **2010**, *21*, 128. [[CrossRef](#)] [[PubMed](#)]
45. Vickers, A.J.; Van Calster, B.; Steyerberg, E.W. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* **2016**, *352*, i6. [[CrossRef](#)]
46. Kerr, K.F.; Brown, M.D.; Zhu, K.; Janes, H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J. Clin. Oncol.* **2016**, *34*, 2534. [[CrossRef](#)] [[PubMed](#)]
47. Baker, S.G. Putting risk prediction in perspective: relative utility curves. *JNCI J. Natl. Cancer Inst.* **2009**, *101*, 1538–1542. [[CrossRef](#)]
48. Good, I.J. Rational decisions. In *Breakthroughs in Statistics*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 365–377.
49. Uno, H.; Cai, T.; Tian, L.; Wei, L. Evaluating prediction rules for t-year survivors with censored regression models. *J. Am. Stat. Assoc.* **2007**, *102*, 527–537. [[CrossRef](#)]
50. Gerdts, T.A.; Schumacher, M. Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biom. J.* **2006**, *48*, 1029–1040. [[CrossRef](#)] [[PubMed](#)]
51. Uno, H.; Cai, T.; Pencina, M.J.; D’Agostino, R.B.; Wei, L. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.* **2011**, *30*, 1105–1117. [[CrossRef](#)] [[PubMed](#)]
52. Guan, Z. A Proper Scoring Rule for Validation of Competing Risks Models. *arXiv* **2021**, arXiv:2104.01000.
53. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [[CrossRef](#)]
54. Parmigiani, G.; Inoue, L. *Decision Theory: Principles and Approaches*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 812.
55. National Comprehensive Cancer Network. Genetic/Familial High-Risk Assessment: Breast and Ovarian (Version 3.2019). 2019. Available online: [https://www2.tri-kobe.org/nccn/guideline/gynecological/english/genetic\\_familial.pdf](https://www2.tri-kobe.org/nccn/guideline/gynecological/english/genetic_familial.pdf) (accessed on 3 May 2020).
56. Metcalfe, K.A.; Birenbaum-Carmeli, D.; Lubinski, J.; Gronwald, J.; Lynch, H.; Moller, P.; Ghadirian, P.; Foulkes, W.D.; Klijn, J.; Friedman, E.; et al. International variation in rates of uptake of preventive options in BRCA1 and BRCA2 mutation carriers. *Int. J. Cancer* **2008**, *122*, 2017–2022. [[CrossRef](#)]
57. Gail, M.H. Performance of BCRAT in high-risk patients with breast cancer. *Lancet Oncol.* **2019**, *20*, e285. [[CrossRef](#)]
58. McCarthy, A.M.; Liu, Y.; Ehsan, S.; Guan, Z.; Liang, J.; Huang, T.; Hughes, K.; Semine, A.; Kontos, D.; Conant, E.; et al. Validation of Breast Cancer Risk Models by Race/Ethnicity, Family History and Molecular Subtypes. *Cancers* **2021**, *14*, 45. [[CrossRef](#)]
59. Solikhah, S.; Nurdjannah, S. Assessment of the risk of developing breast cancer using the Gail model in Asian females: A systematic review. *Heliyon* **2020**, *6*, e03794. [[CrossRef](#)] [[PubMed](#)]
60. Nanda, R.; Schumm, L.P.; Cummings, S.; Fackenthal, J.D.; Sveen, L.; Ademuyiwa, F.; Cobleigh, M.; Esserman, L.; Lindor, N.M.; Neuhausen, S.L.; et al. Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. *JAMA* **2005**, *294*, 1925–1933. [[CrossRef](#)] [[PubMed](#)]
61. Huo, D.; Senie, R.T.; Daly, M.; Buys, S.S.; Cummings, S.; Ogutha, J.; Hope, K.; Olopade, O.I. Prediction of BRCA mutations using the BRCAPRO model in clinic-based African American, Hispanic, and other minority families in the United States. *J. Clin. Oncol.* **2009**, *27*, 1184. [[CrossRef](#)] [[PubMed](#)]
62. Liang, J.W.; Idos, G.E.; Hong, C.; Gruber, S.B.; Parmigiani, G.; Braun, D. Statistical methods for Mendelian models with multiple genes and cancers. *Genet. Epidemiol.* **2022**, *46*, 395–414. [[CrossRef](#)] [[PubMed](#)]
63. Elston, R.C.; Stewart, J. A general model for the genetic analysis of pedigree data. *Hum. Hered.* **1971**, *21*, 523–542. [[CrossRef](#)]
64. Chen, J.; Pee, D.; Ayyagari, R.; Graubard, B.; Schairer, C.; Byrne, C.; Benichou, J.; Gail, M.H. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J. Natl. Cancer Inst.* **2006**, *98*, 1215–1226. [[CrossRef](#)]

65. Kalbfleisch, J.D.; Prentice, R.L. *The Statistical Analysis of Failure Time Data*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 360.
66. Whittemore, A.S.; Gong, G.; Itnyre, J. Prevalence and contribution of BRCA1 mutations in breast cancer and ovarian cancer: results from three US population-based case-control studies of ovarian cancer. *Am. J. Hum. Genet.* **1997**, *60*, 496.
67. Fine, J.P.; Gray, R.J. A proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* **1999**, *94*, 496–509. [[CrossRef](#)]
68. Li, G.; Guo, X.; Tang, L.; Chen, M.; Luo, X.; Peng, L.; Xu, X.; Wang, S.; Xiao, Z.; Yi, W.; et al. Analysis of BRCA1/2 mutation spectrum and prevalence in unselected Chinese breast cancer patients by next-generation sequencing. *J. Cancer Res. Clin. Oncol.* **2017**, *143*, 2011–2024. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.