Recent BRCAPRO upgrades significantly improve calibration

Emanuele Mazzola[1],[*], **Jonathan Chipman**[1],[*], **Su-Chun Cheng**[1], and **Giovanni Parmigiani**[1]

[1]Dana-Farber Cancer Institute; Department of Biostatistics & Computational Biology; 450 Brookline Avenue, Boston 02115, MA, USA

Abstract

The recent release of version 2.0-8 of the BayesMendel package contains an updated BRCAPRO risk prediction model, which includes revised modeling of Contralateral Breast Cancer (CBC) penetrance, provisions for pedigrees of mixed ethnicity and an adjustment for mastectomies among family members. We estimated penetrance functions for contralateral breast cancer by a combination of parametric survival modeling of literature data and deconvolution of SEER9 data. We then validated the resulting updated model of CBC in BRCAPRO by comparing it with the previous release (BayesMendel 2.0-7), using pedigrees from the Cancer Genetics Network (CGN) Model Validation Study. Version 2.0-8 of BRCAPRO discriminates BRCA1/BRCA2 carriers from non-carriers with similar accuracy compared to the previous version (increase in AUC: 0.0043), is slightly more precise in terms of RMSE (decrease in RMSE: 0.0108), and it significantly improves calibration (ratio of observed to expected events of 0.9765 in version 2.0-8, compared to 0.8910 in version 2.0-7). We recommend that the new version be used in clinical counseling, particularly in settings where families with CBC are common.

Keywords

Contralateral Breast cancer; risk assessment; BRCA1; BRCA2; cumulative risk; calibration

Introduction

The BRCA1 and BRCA2 genes help explain about 5–10% of breast cancer cases [1,2]. About 12% of women in the general population will develop breast cancer sometime during their lives [3]. By contrast, according to recent estimates, 55 to 65% of women who inherit a harmful BRCA1 mutation and around 45% of women who inherit a harmful BRCA2 mutation will develop breast cancer by age 70 years [4,5]. Models to predict carrier status are commonly used clinically. Among these, BRCAPRO is the most commonly used, and consistently ranks among the best performing in validation studies [6]. BRCAPRO is continually improved based upon current literature [7–11] and has been incorporated into software used by clinicians, researchers, and software developers [12–15]. The most recent

version of BRCAPRO, included in the BayesMendel 2.0-8 package [16], provides updated estimates for Contralateral Breast Cancer (CBC) penetrance, can handle pedigrees including multiple ethnicities, and adjusts for mastectomies among any of the relatives. CBC occurrence was previously estimated via the simplifying assumption that the first and second diagnoses were independent. Families with CBC had been previously identified as being one of the subgroups where predictions provided by BRCAPRO are less accurate [17]. Recent literature provides further evidence of a strong dependence between fist and second diagnoses [18–20] as well as explicit estimates of the cumulative incidence of CBC among carriers [21]. Before distributing the upgraded model, we validated it using data from the Cancer Genetics Network (CGN) Model Validation Study [6].

Materials and Methods

Validation data

The CGN Model Validation Study comprises pedigrees collected in 8 high-risk counseling clinics that well represent the populations within which the BRCAPRO model is more commonly applied. The study is described in detail elsewhere [6]. For reproducibility, we created an R package that reads in the raw data provided by the 8 centers, performs any necessary preprocessing, and reproduces the BRCAPRO risks evaluated independently at the sites using independent software. Across the eight sites, 2089 pedigrees were provided, of which 2038 run without error through both versions of BRCAPRO, and were thus eligible for a head-to-head model comparison. Since we do not have separate information on ethnicity of individual family members, nor information on mastectomies for these individuals, we focus on improvements resulting from the updated CBC penetrances. We compare the ability of the models to a) discriminate carriers from non-carriers using the area under the ROC curve (AUC [22, 23]), b) increase precision of estimated carrier probabilities using the root mean square error of prediction (for MSE see [22, 23]), and c) estimate the overall number of observed carriers using the observed to expected ratio (OE [22, 23]). For each metric, we evaluate the estimates, their 95% bias-corrected accelerated (BCa) bootstrap confidence intervals, the estimates of the differences between the two versions of BRCAPRO, and their corresponding 95% BCa bootstrap confidence intervals. We assessed overall performance among the 2038 pedigrees and considered the following two subgroups: 322 families with a CBC diagnosis in any of the relatives (subgroup 1); 155 families in which the proband is diagnosed with CBC (subgroup 2). We used the R package ROCR [24] to estimate the AUC and MSE.

Statistical modeling

We estimated the penetrances of Contralateral Breast Cancer (CBC) for carriers of mutations in either one of the BRCA genes or both, and non-carriers as functions. For consistency with earlier versions of BRCAPRO we used a discrete time variable t from 0 to 110, interpreted as difference between the ages of the two breast cancer diagnoses, in years. A value of $t=0$ indicates that the first unilateral and second contralateral diagnoses occurred within a year of each other. We use the term unilateral only for the first diagnosis; the second is always called contralateral. We estimated penetrances for BRCA1 and BRCA2 carriers from published cumulative risks estimates [21]. To estimate non-carrier penetrance,

we first estimated the general population penetrance from SEER9 database and subtracted the penetrance attributable to carriers.

Penetrance of BRCA Carriers

Graeser et al. [21], estimated the cumulative risk of CBC following a first unilateral breast cancer diagnosis in relatives of BRCA1 or BRCA2 carriers. This was a retrospective, multi-center study based on the German Consortium for Hereditary Breast and Ovarian Cancer, ranging from 1996 until 2008, and including a total of 1,520 families with a deleterious germline mutation in either BRCA1 or BRCA2. Exclusion criteria from the study were: 1. relatives testing negative for the known mutation in the family; 2. synchronous bilateral or noninvasive breast cancer; 3. insufficient information regarding age at cancer events or bilateral mastectomy. The final study cohort comprised 2,020 women with unilateral breast cancer (978 index patients and 1,042 relatives); the group of index patients were not used to estimate risk of CBC following first diagnosis because they were already selectively chosen to be DNA tested. The authors provided the cumulative risk of CBC at 5, 10, 15, and 25 years following the original diagnosis. In relatives of BRCA1 carriers, age of unilateral breast cancer diagnosis was associated with risk of CBC, and estimates were provided for women diagnosed with unilateral breast cancer before age 40, between 40 and 49, and older. There does not appear to be a pronounced difference in cumulative risk between women diagnosed between 40 to 49 years old and women diagnosed later.

We used maximum-likelihood estimation, assuming an underlying Weibull distribution, to fit these cumulative risk estimates. For BRCA1, we fit two separate models: one for women diagnosed with unilateral breast cancer before age 40 and a second model for the two older age categories combined, using their averaged cumulative risks as input. For BRCA2, differences among cumulative risks for any the age categories failed to achieve significance [21]; we thus we fit a single cumulative risks model to all age groups combined.

BRCAPRO needs to cover a broad range of combinations of ages of first and second diagnoses, including some that are so far apart to have never been observed in [21] or SEER. To cover these unlikely cases we extrapolated the CBC penetrance curves beyond the 25 year follow-up covered by [21]. To this end, we made the assumption that the penetrance follows an exponential decay starting at year 26 after the first unilateral diagnoses, and ending with a penetrance of essentially zero at year 110.

We also need to cover the unlikely scenario of carriers of carriers of both a BRCA1 and BRCA2 mutation. In keeping with earlier versions of BRCAPRO we assume that the age of onset for these individuals is distributed as the minimum of the corresponding random variables for carriers of a single mutation. The cumulative risk $R_{12}(t)$ for carriers of both mutations is thus obtained from the BRCA1 cumulative risk $R_1(t)$, and the BRCA2 cumulative risk $R_2(t)$ as

$$R_{12}(t) = 1 - [1 - R_1(t)][1 - R_2(t)], \quad t = 0, \ldots .110. \quad (1)$$

Penetrance of the General Population

From the SEER9 database, covering diagnoses from 1973 to 2006, we extracted patients who experienced invasive CBC. Repeated diagnoses, bilateral diagnoses, and diagnoses with unknown laterality were removed for a total available sample of 457,304 women aged 11 to 108 at unilateral breast cancer diagnosis. We split the general population into two different age groups, containing respectively 29,659 women aged <40 and 427,645 women aged 40 at their first breast cancer diagnosis. We modeled the time to CBC from unilateral breast cancer diagnosis using a Weibull parametric survival curve and derived the cumulative risk. The maximum time to CBC was 34 years after unilateral breast cancer diagnosis.

Penetrance for Non-Carriers

We estimated the cumulative risk for non-carriers $R_{noncar}(t)$ from the cumulative risk of carriers $R_{car}(t)$ and the cumulative risk for the general population $R_{pop}(t)$. Let $\theta_1(t)$, $\theta_2(t)$ be the allele frequencies of deleterious mutations in the BRCA1 and BRCA2 genes respectively. Also let $\theta_{12}(t)$ be the allele frequency of a mutation in both BRCA1 and BRCA2. We can express $R_{pop}(t)$ as a mixture of $R_{car}(t)$ and $R_{noncar}(t)$

$$R_{pop}(t) = \pi_{car} R_{car}(t) + \pi_{noncar} R_{noncar}(t), \quad t = 0, \ldots, 110, \quad (2)$$

where $\pi_{car} = \theta_1 + \theta_2 - \theta_{12}$ and $\pi_{noncar} = 1 - \pi_{car}$ are respectively the probabilities of a being carrier and a non-carrier in the general population. The term $R_{car}(t)$ in (2) can be written as a linear combination:

$$R_{car}(t) = \theta_1 R_1(t) + \theta_2 R_2(t) - \theta_{12} R_{12}(t)$$

where $R_1(t)$, $R_2(t)$, $R_{12}(t)$ are defined as above. By solving equation (2) for $R_{noncar}(t)$, we obtained the cumulative risk for noncarriers. As before, for years 35–110 following unilateral breast cancer diagnosis, we assumed that the penetrance function follows an exponential decay to essentially zero, as previously explained.

Further Adjustments

We applied some minor additional adjustments to the curves. Because the first CBC cumulative risk estimates in [21] was 5 years after unilateral breast cancer diagnosis, we assumed the cumulative risks $R_1(t)$ and $R_2(t)$ for BRCA carriers to be linear between year 1 and year 5; moreover, we set $R_1(0)$ equal to $R_1(1)$ and $R_2(0)$ equal to $R_2(1)$ respectively for BRCA1 and BRCA2 mutation carriers. This removed a singularity at time $t = 0$ given by the Weibull parametric model, which would have made the estimated probability of a contemporaneous diagnosis of contralateral breast cancer greater than one. We also removed a singularity at time $t = 0$ for the general and non-carrier population penetrance curves, assuming a linear cumulative risk between times $t = 0$ and $t = 1$. Figure 1 shows the final penetrance density functions that have been included in the current implementation of BRCAPRO 2.0-8.

Results

Performance of BRCAPRO 2.0-8

As expected, only probands in subgroup 1 have a modified risk of being a BRCA carrier in BRCAPRO 2.0-8 compared to 2.0-7. Figure 2 provides an overall comparison. For the vast majority of families with CBC, the carrier probability is reduced in the new version. This is because, generally, two positively correlated diagnoses provide less evidence towards increased risk than would two independent diagnoses. A large number of families, highly enriched for non-carriers moves from high to low risk by the typical definitions of risk used clinically (e.g. 5% or 10%). Figure 3 further breaks down CBC families depending on whether the proband or a relative is affected with CBC (panel a), and depending on the time interval between the two diagnoses (panel b). The carrier risk decreased more pronouncedly if the CBC occurred in the proband and/or if fewer years passed between unilateral and contralateral breast diagnoses. While in most families with CBC, the estimated carrier risk is lower in the revised model, exceptions occur when at least 12 years passed between diagnoses.

The two versions of BRCAPRO discriminate similarly well between carriers and non-carriers overall (difference in AUC between release 2.0-8 and release 2.0-7 =0.0043), in subgroup 1 (difference in AUC = 0.0002) and in subgroup 2 (difference in AUC = 0.0068); see Table 1 for the BCa 95% confidence intervals. The new version has increased precision as measured by a statistically significant decrease in RMSE of 0.0108 (c.i. −0.0154 to −0.0067) (see also Table 1). As expected, this trend in RMSE is driven by families in subgroup 1 presenting with a statistically significant decrease in RMSE of 0.0551 (c.i. −0.0761 to −0.0347), and in subgroup 2 with a statistically significant decrease in RMSE of 0.0633 (c.i. −0.0984 to −0.0306).

The calibration of BRCAPRO improves in version 2.0-8. The new OE of 0.98, a statistically significant increase of 0.09 with respect to version 2.0-7, and is closer to the target value of 1; when this metric is considered separately for the two genes, the OE for BRCA1(2) carrier status is 1.04 (0.89). In both subgroups 1 and 2, this is an improvement (0.73 for version 2.0-8 from 0.55 for version 2.0-7, and 0.8 from 0.58 respectively). In BRCAPRO 2.0-8, the five-year risk of a CBC diagnosis for probands with an unilateral breast diagnosis is lower than in the previous version. The biggest impacts upon five-year risk of CBC occur when the first diagnosis is before age 40, and in presence of family CBC (Figure 4). For unaffected probands, the risk of a breast cancer diagnosis within five years remains the same if there is no family history of CBC, but decreases otherwise (Figure 5).

Discussion

We updated the BRCAPRO and improved the way in which it handles CBC. These changes mitigate a previous problem with overestimation of carrier risk in families with CBC, and lead to significantly improved model calibration, as well as a significant improvement in estimation accuracy, of the order of twenty or more percentage points, in these families. Based on this evaluation, we suggest that the new version should be used in clinical

counseling, particularly in settings where families with CBC are common and where affected probands are counseled about their contralateral recurrence risk.

Acknowledgments

References

1. American Cancer Society. Cancer Facts & Figures 2013. American Cancer Society; 2013.

2. National Cancer Institute. Genetics of breast and ovarian cancer (PDQ). 2013. http://www.cancer.gov/cancertopics/pdq/genetics/breast-and-ovarian/HealthProfessional/page2

3. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: A combined analysis of 22 studies. Am J Hum Genet. 2003; 72:1117–1130. [PubMed: 12677558]

4. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol. 2007; 25:1329–1333. [PubMed: 17416853]

5. Howlader, N.; Noone, AM.; Krapcho, M.; Garshell, J.; Neyman, N.; Altekruse, SF., et al. SEER Cancer Statistics Review, 1975–2010. National Cancer Institute; Bethesda, MD: Apr. 2013 http://seer.cancer.gov/csr/1975_2010/, based on November 2012 SEER data submission, posted to the SEER web site

6. Parmigiani G, Chen S, Iversen ES Jr, Friebel TM, Finkelstein DM, Anton-Culver H, et al. Validity of Models for Predicting BRCA1 and BRCA2 Mutations. Ann Intern Med. 2007; 147:441–450. [PubMed: 17909205]

7. Katki HA. Incorporating medical interventions into carrier probability estimation for genetic counseling. BMC Med Genet. 2007; 8:13. [PubMed: 17378937]

8. Tai YC, Domchek S, Parmigiani G, Chen S. Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. J Natl Cancer Inst. 2007; 99:1811–1814. [PubMed: 18042939]

9. Tai YC, Chen S, Parmigiani G, Klein AP. Incorporating tumor immunohistochemical markers in BRCA1 and BRCA2 carrier prediction. Breast Cancer Res. 2008; 10:401. [PubMed: 18466632]

10. Chen S, Blackford A, Parmigiani G. Tailoring BRCAPRO to Asian-Americans. J Clin Oncol. 2009; 27:642–643. [PubMed: 19075251]

11. Biswas S, Tankhiwale N, Blackford A, Barrera AM, Ready K, Lu K, et al. Assessing the added value of breast tumor markers in genetic risk prediction model BRCAPRO. Breast Cancer Res Treat. 2012 May.133:347–55. [PubMed: 22270937]

12. CaGene. http://www4.utsouthwestern.edu/breasthealth/cagene/

13. HughesRiskApps. http://www.hughesriskapps.com/

14. Ozanne EM, Loberg A, Hughes S, Lawrence C, Drohan B, Semine A, et al. Identification and management of women at high risk for hereditary breast/ovarian cancer syndrome. Breast J. 2009; 15:155–62. [PubMed: 19292801]

15. Chipman J, Drohan B, Blackford A, Parmigiani G, Hughes K, Bosinoff P. Providing access to risk prediction tools via the HL7 XML-formatted risk web service. Breast Cancer Res Treat. 2013 Jul. 140:187–93. [PubMed: 23793601]

16. Chen S, Wang W, Broman KW, Katki HA, Parmigiani G. BayesMendel: and R environment for Mendelian risk prediction. Stat Appl Genet Mol Biol. 2004; 3:Article 21.

17. Ready KJ, Vogel KJ, Atchley DP, Broglio KR, Solomon KK, Amos C. Accuracy of the BRCAPRO model among women with bilateral breast cancer. Cancer. 2009; 115:725–730. [PubMed: 19127556]

18. Chen Y, Thompson W, Semenciw R, Mao Y. Epidemiology of contralateral breast cancer. Cancer Epidemiol Biomarkers Prev. 1999; 8:855–861. [PubMed: 10548312]

19. Metcalfe K, Gershman S, Ghadirian P, Lynch HT, Snyder C, Tung N, et al. Contralateral breast cancer in BRCA1 and BRCA2 mutation carriers. J Clin Oncol. 2004; 15:2328–2335. [PubMed: 15197194]

20. Rubino C, Arriagada R, Delaloge S, Lê MG. Relation of risk of contralateral breast cancer to the interval since the first primary tumour. Br J Cancer. 2010; 102:213–219. [PubMed: 19920826]

21. Graeser MK, Engel C, Rhiem K, Gadzicki D, Bick U, Kast K, et al. Contralateral Breast Cancer Risk in BRCA1 and BRCA2 Mutation Carriers. J Clin Oncol. 2009; 27:5887–5892. [PubMed: 19858402]

22. Pencina MJ, D'Agostino RB, Pencina KM, Janssens ACJW, Greenland P. Interpreting incremental value of markers added to risk prediction models. Am J Epidemiol. 2012; 176:473–481. [PubMed: 22875755]

23. Kerr KF, Bansal A, Pepe M. Further insight into the incremental value of new markers: the interpretation of performance measures and the importance of clinical context. Am J Epidemiol. 2012; 176:482–487. [PubMed: 22875756]

24. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005; 21:3940–3941. [PubMed: 16096348]
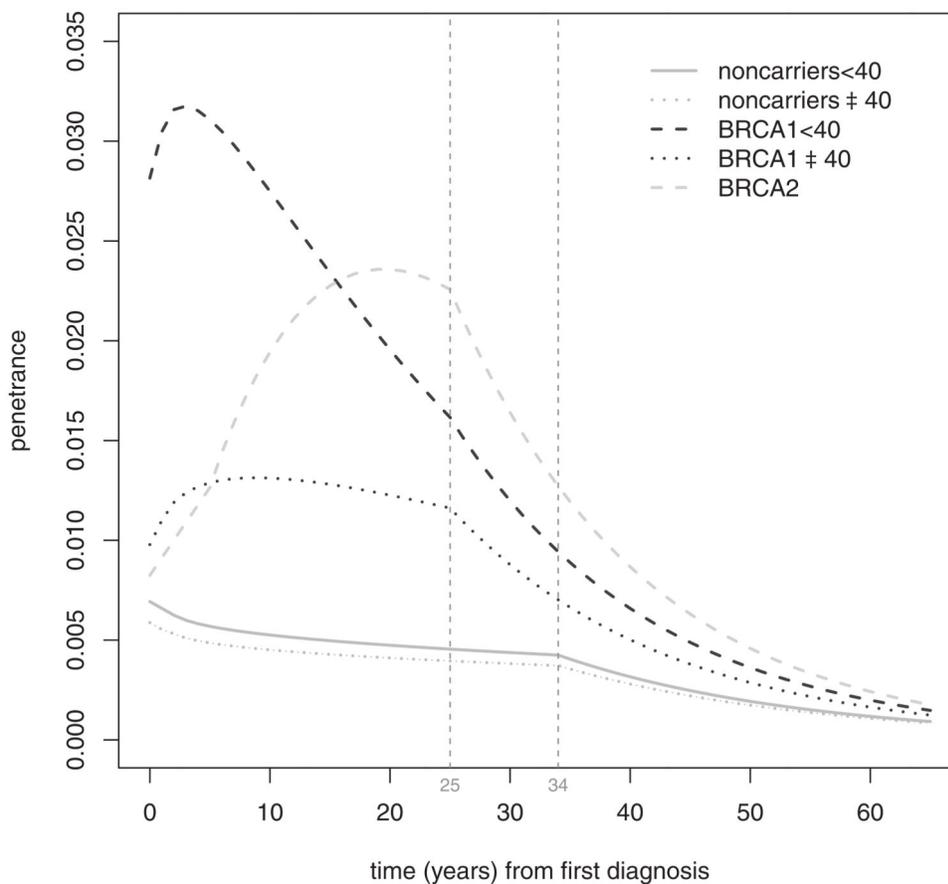
Figure 1.
Smoothed, age-stratified penetrance density curves for carriers of either a BRCA1 or a BRCA2 mutation, and the non-carrier population. Vertical lines at 25 and 34 years after the first diagnosis of breast cancer, indicate the last available piece of data in [21] and in the SEER database, respectively. After these limits, the corresponding penetrance density functions are assumed to decrease exponentially.
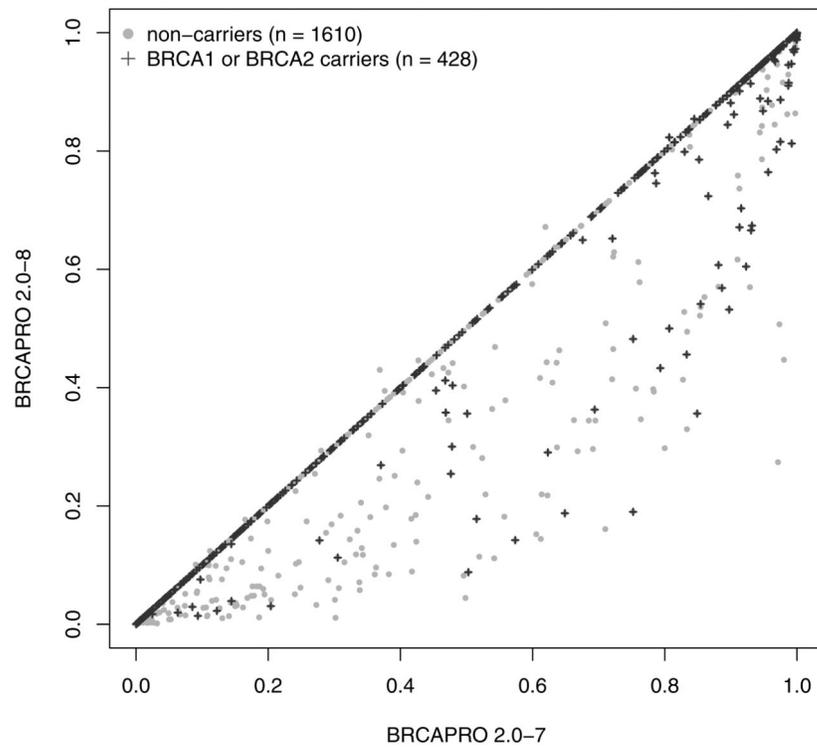
Figure 2.
Comparison of the predicted risk of carrying a BRCA1 or BRCA2 mutation between BRCAPRO 2.0-7 and 2.0-8. Patients who don't carry a mutation are indicated by the grey dots. Patients who tested positive as BRCA mutation carriers are represented by the black crosses; risk predictions above the diagonal are elucidated in Figure 3b.

Figure 3a
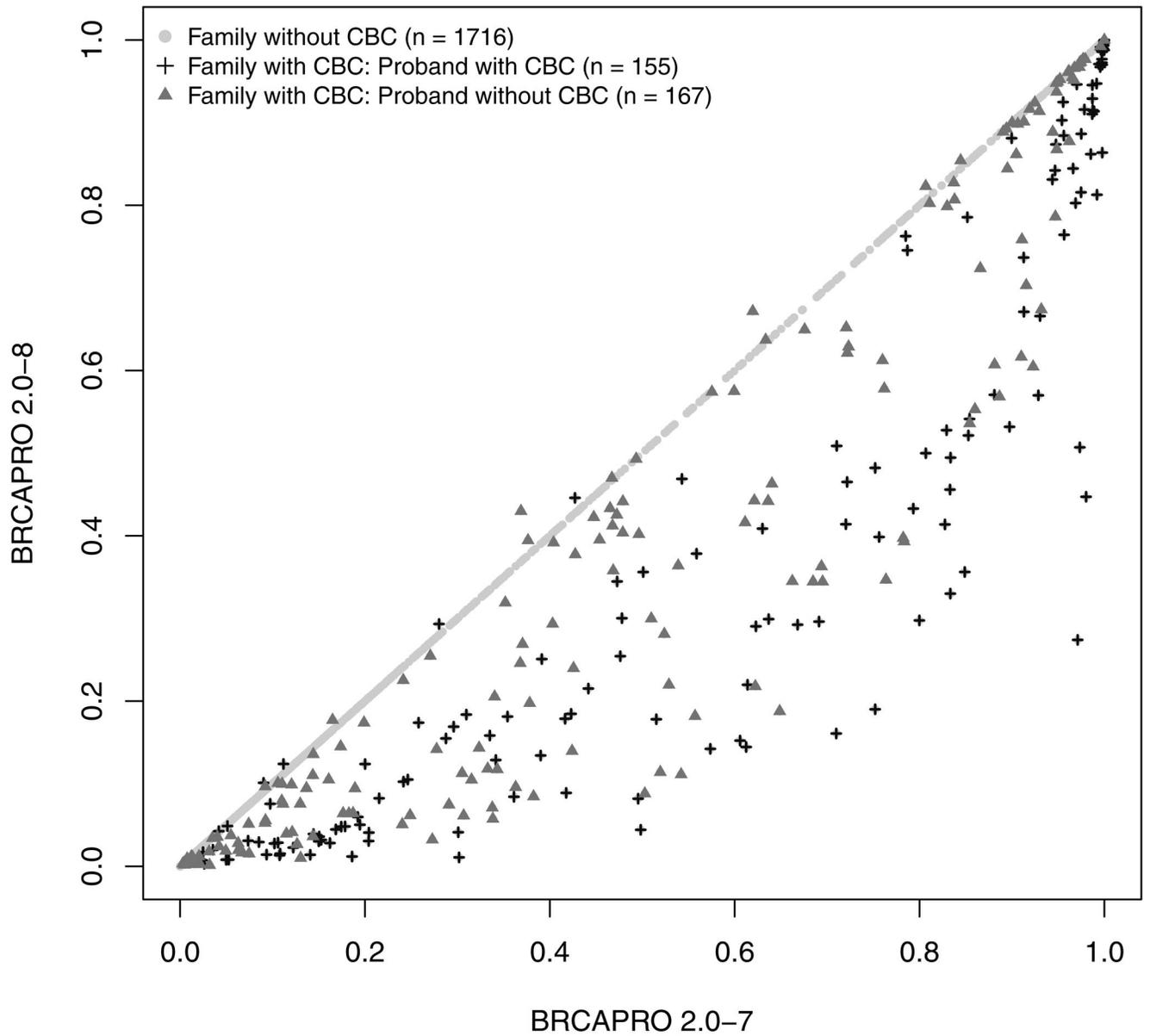
Figure 3b
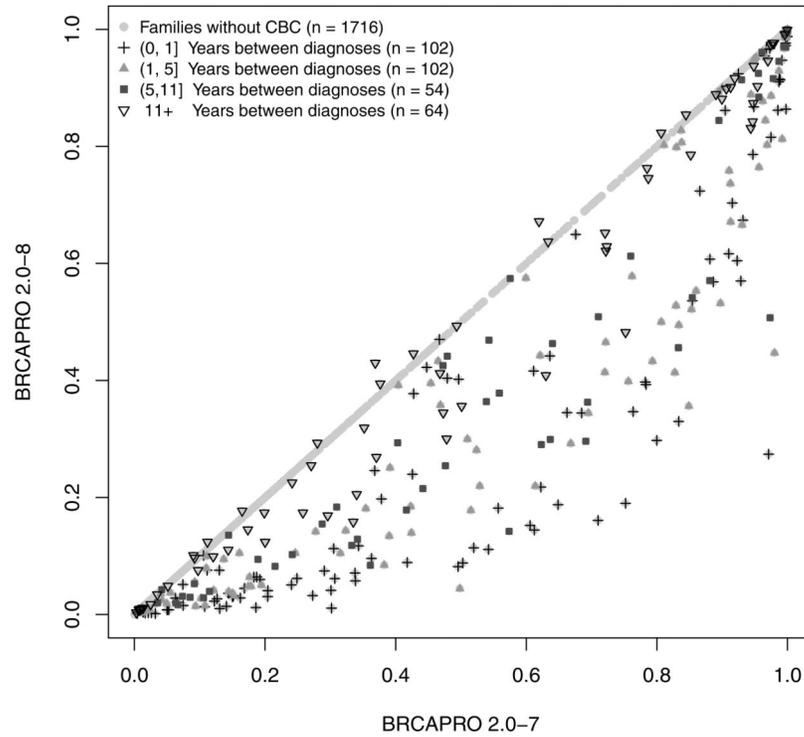
Figure 3.

This figure shows two stratifications of the information reported in Figure 2; (a): comparison of risk prediction of being a BRCA carrier between BRCAPRO 2.0-7 and 2.0-8; 322 families have members affected with CBC; in 155 of these the proband is affected with CBC; (b) comparison of risk prediction of being a BRCA carrier between BRCAPRO 2.0-7 and 2.0-8, stratified by difference in years from the first diagnosis of breast cancer. In families with multiple incidences of CBC, the average number of years between diagnoses is reported.
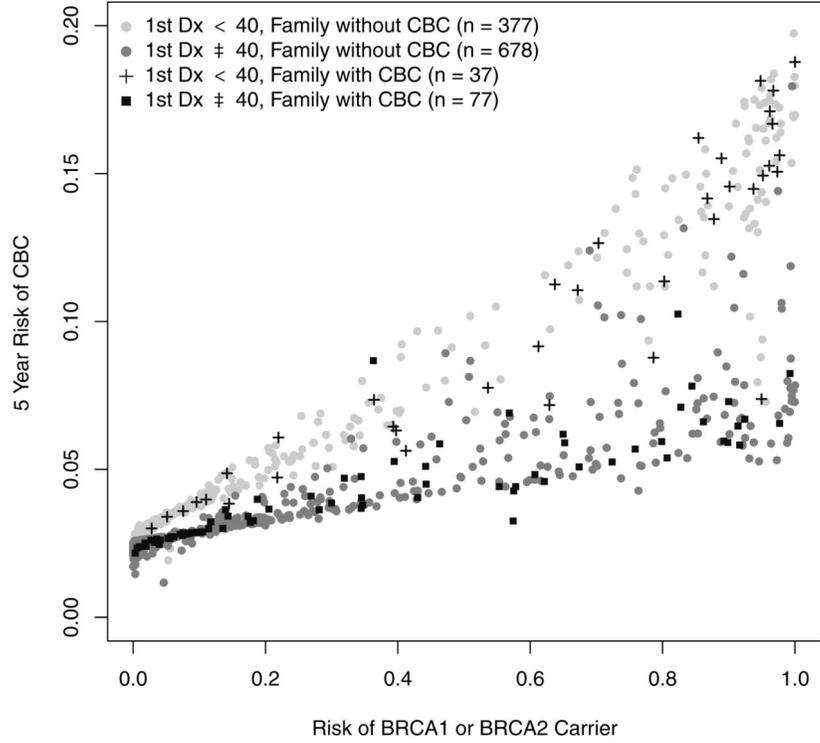
Figure 4.
BRCAPRO 2.0-8 risk of a diagnosis of CBC within 5 years of the first diagnosis of breast cancer as a function of the risk of carrying a mutation in either one of the BRCA genes, stratified both by age at first diagnosis (less than 40, and greater than or equal to 40), and by the presence or absence of CBC in the family.
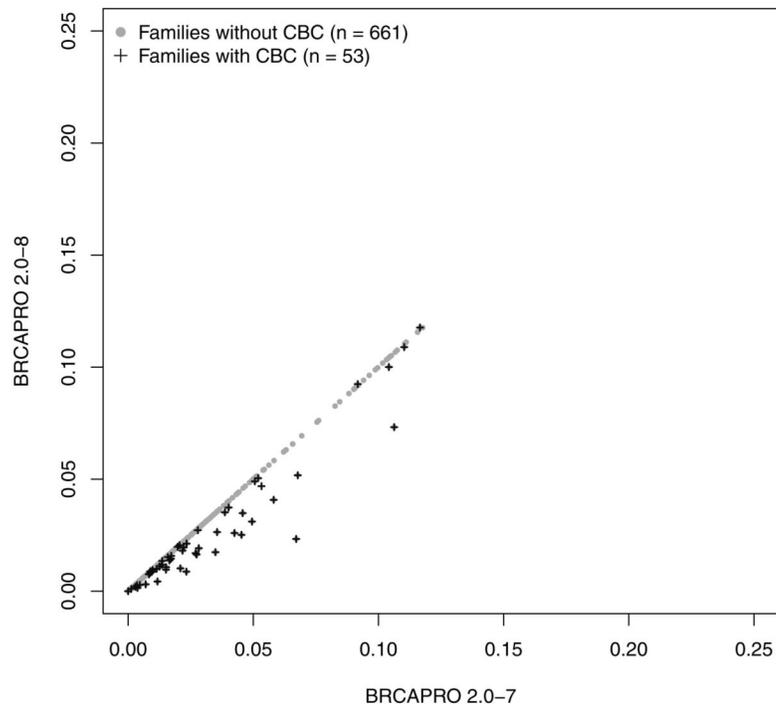
Figure 5.
Comparison of risk predictions for breast cancer within 5 years for undiagnosed probands, both for patients exhibiting (black crosses) or not exhibiting (grey dots) a family history of CBC.

Table 1

Comparison of performance of BRCAPRO version 2.0-7 and version of 2.0-8 with 95% BCa marginal confidence intervals. Rows labeled by ∆ contain the difference of the figure of merit between BRCAPRO 2.0-7 and 2.0-8, with corresponding 95% confidence intervals. The differences ∆OE and ∆RMSE are statistically significant.

	Overall	CBC in Family	CBC Proband
AUC 2.0-7	0.7884 (0.7613, 0.8126)	0.7785 (0.7177, 0.8295)	0.7479 (0.6533, 0.8220)
AUC 2.0-8	0.7927 (0.7661, 0.8169)	0.7788 (0.7185, 0.8311)	0.7547 (0.6623, 0.8274)
∆AUC	0.0043 (0.0005, 0.0082)	0.0002 (−0.0189, 0.0194)	0.0068 (−0.0016, 0.0312)
OE 2.0-7	0.8910 (0.8239, 0.9608)	0.5515 (0.4700, 0.6360)	0.5837 (0.4680, 0.7051)
OE 2.0-8	0.9765 (0.9036, 1.0519)	0.7296 (0.6198, 0.8402)	0.8018 (0.6445, 0.9750)
∆OE	0.0855 (0.0722, 0.1021)	0.1781 (0.1441, 0.2186)	0.2181 (0.1634, 0.2894)
RMSE 2.0-7	0.3854 (0.3698, 0.4020)	0.4998 (0.4652, 0.5355)	0.5314 (0.4789, 0.5848)
RMSE 2.0-8	0.3745 (0.3587, 0.3910)	0.4448 (0.4090, 0.4805)	0.4682 (0.4162, 0.5240)
∆RMSE	−0.0108 (−0.0154, −0.0067)	−0.0551 (−0.0761, −0.0347)	−0.0633 (−0.0984, −0.0306)