




Validation of a personalized risk prediction model for contralateral breast cancer

Marzana Chowdhury¹ · David Euhus² · Banu Arun³ · Chris Umbricht² · Swati Biswas¹  · Pankaj Choudhary¹

Received: 27 February 2018 / Accepted: 17 March 2018 / Published online: 24 March 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Purpose Women diagnosed with unilateral breast cancer are increasingly choosing to remove their other unaffected breast through contralateral prophylactic mastectomy (CPM) to reduce the risk of contralateral breast cancer (CBC). Yet a large proportion of CPMs are believed to be medically unnecessary. Thus, there is a pressing need to educate patients effectively on their CBC risk. We had earlier developed a CBC risk prediction model called CBCRisk based on eight personal risk factors.

Methods In this study, we validate CBCRisk on independent clinical data from the Johns Hopkins University (JH) and MD Anderson Cancer Center (MDA). Women whose first breast cancer diagnosis was either invasive and/or ductal carcinoma in situ and whose age at first diagnosis was between 18 and 88 years were included in the cohorts because CBCRisk was developed specifically for these women. A woman who develops CBC is called a case whereas a woman who does not is called a control. The cohort sizes are 6035 (with 117 CBC cases) for JH and 5185 (with 111 CBC cases) for MDA. We computed the relevant calibration and validation measures for 3- and 5-year risk predictions.

Results We found that the model performs reasonably well for both cohorts. In particular, area under the receiver-operating characteristic curve for the two cohorts range from 0.61 to 0.65.

Conclusions With this independent validation, CBCRisk can be used confidently in clinical settings for counseling BC patients by providing their individualized CBC risk. In turn, this may potentially help alleviate the rate of medically unnecessary CPMs.

Keywords CBCRisk · Contralateral breast cancer · Contralateral prophylactic mastectomy · Absolute risk

Introduction

The rates of contralateral prophylactic mastectomy (CPM) among patients diagnosed with unilateral breast cancer (BC) have been increasing sharply over the last two decades [1–4]. However, most of these CPMs are believed to be medically unnecessary because the risk of contralateral breast cancer

(CBC) has been, in fact, falling mostly due to increased use of effective adjuvant therapies such as tamoxifen and aromatase inhibitors for treating the first primary breast cancer [5]. While unilateral BC patients carrying BRCA mutation or having family history of BC indeed have a high risk of developing CBC, a majority of women have much lower risk of developing CBC. Even more perplexing is the fact that the rate of CPM has increased the most in a group of BC patients who are at the least risk of getting CBC [6].

These paradoxical observations have baffled the scientific community leading to a flurry of investigations to uncover the reasons driving the decision to undergo CPM. Although this decision-making process is complex with several factors playing roles in varying degrees for different women, research shows that many patients tend to overestimate their CBC risk, sometimes even by a factor of 5–10 [6–8]. Thus, it is natural for them to instinctively lean towards CPM if only as an anxiety-relieving tool. Although CPM does reduce the risk of CBC, there is no convincing evidence that it prolongs survival

✉ Swati Biswas
swati.biswas@utdallas.edu

✉ Pankaj Choudhary
pankaj@utdallas.edu

¹ Department of Mathematical Sciences, University of Texas at Dallas, 800 W Campbell Rd., FO 35, Richardson, TX 75080, USA

² Division of Surgical Oncology, Johns Hopkins University, Baltimore, USA

³ Department of Breast Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, USA

[6]. In fact, for majority of patients, their CBC risk is much smaller than the risk of recurrence from their index cancer [3]. Moreover, CPM also has a substantial number of side effects and can have an adverse effect on a woman's physical and psychological health and well-being [9]. Thus, the decision-making process for CPM should weigh these facts along with the loss of a healthy breast in light of the actual risk of CBC.

These findings call for a pressing need to educate patients effectively on their CBC risk. For this task, physicians need a statistical model for risk prediction of CBC based on patient's personal risk factors. We had earlier developed such a model named CBCRisk for predicting absolute risk of CBC for women diagnosed with invasive BC or ductal carcinoma in situ (DCIS) in one breast [10]. The model was built using US population-based datasets from Breast Cancer Surveillance Consortium (BCSC) [11] and Surveillance, Epidemiology, and End Results (SEER) [12]. It uses eight risk factors from the time of the first BC diagnosis, namely, age at first BC diagnosis, anti-estrogen therapy, family history of BC, high-risk pre-neoplasia status, breast density, estrogen receptor (ER) status, type of first BC, and age at first birth. The directions of association of these factors with the risk of CBC in CBCRisk are the following. Higher risk is associated with *younger* age at first BC diagnosis, *no* anti-estrogen therapy, *presence of* family history of BC, *presence of* high-risk pre-neoplasia, *denser* breast density, *negative* estrogen receptor (ER) status, *presence of DCIS* in first BC, and *older* age at first birth. Except for age at first diagnosis and first BC type, an unknown category is available for all other factors. This way the model can be applied even for women with missing values on some factors with the usual caveats of unknown category, e.g., it is heterogeneous with possibly different compositions across different cohorts. For a woman with cancer in one breast, CBCRisk takes information on these risk factors and her current age as inputs, and provides the absolute risk of her getting CBC within a specified number of years as output. CBCRisk has been implemented in an R software package as well as a web-based app [13].

Here we conduct a study to validate CBCRisk on independent clinical data from the Johns Hopkins University (JH) and the University of Texas MD Anderson Cancer Center (MDA). We primarily focus on absolute risk of developing CBC within three and five years of first BC diagnosis. We could not consider risk estimates for longer periods, e.g., 10 years, because only a small portion of the validation cohorts have been followed up that long.

Data sources

The validation data were from two hospital-based cohorts—JH and MDA. Women whose first BC diagnosis was histologically confirmed as either invasive and/or DCIS and

whose age at first diagnosis was between 18 and 88 years were included in the cohorts because CBCRisk was developed specifically for these women [10]. The follow-up for each woman begins at the time of her first BC diagnosis. Women who underwent CPM during the follow-up period or who got CBC within 6 months of their first BC diagnosis were excluded. Same as the first primary diagnosis, we also require the CBC diagnosis to be histologically confirmed as either invasive and/or DCIS. A woman who develops CBC is called a case whereas a woman who does not is called a control. For a case, the follow-up ends at the time of her CBC diagnosis, whereas for a control, the follow-up ends at censoring or death.

The initial JH cohort consisted of 6447 women whose first BC was diagnosed between 1990 and 2016. After applying the exclusion criteria, 6035 women remained in the cohort consisting of 117 cases and 5918 controls. Of these, there were only 231 women, including 81 cases, who had all eight variables (used by CBCRisk) recorded. For the remaining women, only the age at first diagnosis, first BC type, high-risk pre-neoplasia status, anti-estrogen therapy, and ER status were available. For an unreported variable, the value was set to be unknown as allowed by CBCRisk for most variables. The initial MDA cohort had 6123 women who got first BC diagnosis between 1983 and 2014. After excluding women based on the exclusion criteria, 5185 women remained in the cohort, including 111 cases and 5074 controls. High-risk pre-neoplasia status and breast density were not available in these data, and thus were set as unknown in CBCRisk. Further, age at first birth was also missing for almost 94% of women in this cohort. Duration of follow-up was available for all women in the two cohorts.

The follow-up profiles of the two cohorts are somewhat different. Specifically, the mean and the median follow-up times in the JH cohort are, respectively, 6.4 and 6 years for cases and 8 and 7.4 years for controls, and about 8% of women are followed up for less than one year, whereas the mean and the median follow-up times in the MDA cohort are, respectively, 5.1 and 4.2 years for cases and 5.8 and 4.8 years for controls, and about 1.7% of women are followed up for less than 1 year.

For validation of risk prediction in a specific period, for example, 3 or 5 years after first BC diagnosis, we need the true disease (CBC) status for each woman at the end of that specific risk prediction period. Although this can be determined unambiguously for a woman who ends up developing CBC during her follow-up period (a case), for a control who died or was censored *before* the end of the risk prediction period, her true CBC status remains unknown. Therefore, such women need to be dropped for the validation in that specific risk prediction period. This means that for validation of 3 (or 5 year) risk, we consider only those controls who have been followed up for at least 3 years (or 5 years). This

necessarily implies that the size of the validation cohort for 5-year risk is a subset of that for 3-year risk. Specifically, in the JH cohort, 5026 and 4142 controls have been followed up for at least 3 and 5 years, respectively. The corresponding numbers for the MDA cohort are 3547 and 2478 controls. Moreover, for a case, if her length of follow-up, i.e., the time between her first BC and CBC diagnosis, exceeds 3 years, she, in fact, serves as a control for 3-year risk validation and likewise for 5-year risk validation. Lastly, CBCRisk sets 89 years as the upper age limit for risk calculation (this allows 1-year risk prediction for a woman of 88 years, the oldest age for which CBCRisk can be used). Women whose age at first BC diagnosis is such that they would go over the 89-year limit in 3 years (5 years) are dropped from the cohort for the calculation of 3-year (5-year) risk.

Statistical methods

To validate CBCRisk, we evaluate its calibration and discrimination accuracy in predicting absolute risk of developing CBC over a specified period starting from the age at first BC diagnosis. Two risk prediction periods of 3 and 5 years are considered. We could not consider longer risk prediction periods such as 10 years as only few women have been followed up for that long (as is now clear from the previous section). We begin by computing the absolute risks for each woman in the validation cohorts. We set the current age for each woman in CBCRisk as her age at first BC diagnosis as this is when most women are likely to seek risk assessment for CBC. The calibration is evaluated by comparing the expected number of CBC cases (E) as predicted by the model with the observed number of CBC cases (O) at the end of a risk prediction period through E/O ratio for the entire cohort as well as for subgroups formed by quintiles of predicted risk and by various levels of the risk factors used in CBCRisk. For a given group of women, E is computed by summing the absolute risks of all women in that group. We also compute 95% confidence interval (CI) for E/O based on a normal approximation for natural logarithm of O given by $(E/O) \times \exp(\pm 1.96\sqrt{1/O})$ [14].

To evaluate the discrimination ability, we calculate sensitivity (true positive rate) and specificity (true negative rate) values at various cutoffs and plot sensitivity vs one minus specificity as a receiver-operating characteristic (ROC) curve. The area under the ROC curve (AUC) is then used to summarize the overall discrimination ability of the model. This index ranges between 0 and 1, with 0.5 corresponding to a fair coin toss and 1 corresponding to perfect prediction. The AUC can be interpreted as the probability that the absolute risk of a randomly selected diseased woman is greater than the absolute risk of a randomly selected nondiseased

woman. We also compute a 95% CI for AUC using bootstrap method [15, 16].

We used the statistical software package R for all the computations [17]. The CBCRisk package [13] has been used to calculate the absolute risks based on the CBCRisk model. The package pROC has been used to compute AUC and its 95% bootstrap CI and to draw the ROC curve [18].

Results

Table 1 presents the case–control distributions for the two cohorts together with those of BCSC and SEER data on which CBCRisk was built [10]. The proportion of cases in both the cohorts are smaller than those in BCSC and SEER. Specifically, JH and MDA, respectively, have 0.5 and 0.3% fewer cases (as percentages of respective total cohort sizes) than SEER. Table 2 presents the case–control distributions of JH and MDA cohorts by stratifying them on the eight risk factors used in CBCRisk. Clearly, the compositions of the two cohorts are quite different. Only the distributions according to age at first diagnosis are comparable. Although the distributions according to high-risk pre-neoplasia status, breast density, and age at first birth may also seem somewhat similar but this is primarily driven by the large number of women in the “unknown” category. A majority of women in both cohorts are ER positive but their percentages are not similar. A marked difference exists in distributions according to anti-estrogen therapy, family history, and type of first BC. For example, both anti-estrogen therapy and family history are “unknown” for a majority of JH women, whereas they are, respectively, “yes” and “no” for a majority of MDA women. Likewise, the MDA data has over 98% “pure invasive” as first BC type, in contrast with about 40% “pure invasive,” 20% “pure DCIS,” and 40% “mixed invasive/DCIS” in JH data. This may be because the associated DCIS information may not have been captured in the MDA data in addition to the primary BC type. Thus, a substantial proportion of pure invasive patients could, in fact, be of mixed type. Now we present the validation results. First, we consider the discrimination performance of CBCRisk.

Table 1 Case–control distributions in cohorts

Dataset	Cases	Controls	Cohort size
Johns Hopkins	117 (1.9)	5918 (98.1)	6035
MD Anderson	111 (2.1)	5074 (97.9)	5185
BCSC	1921 (2.5)	75,825 (97.5)	77,746
SEER	19,835 (2.4)	804,933 (97.6)	824,768

The numbers within parentheses indicate percentage of the cohort size

Table 2 Case–control distributions according to risk factors

Risk factor	Johns Hopkins (<i>n</i> = 6035)		MD Anderson (<i>n</i> = 5185)	
	Case	Control	Case	Control
Age at first diagnosis				
< 30	0	55 (0.9)	2 (0.04)	56 (1.1)
30–40	11 (0.2)	500 (8.3)	6 (0.1)	411 (7.9)
40 +	106 (1.8)	5363 (88.9)	103 (2)	4607 (88.9)
Anti-estrogen therapy				
Yes	71 (1.2)	1119 (18.5)	71 (1.4)	3547 (68.4)
No	43 (0.7)	949 (15.7)	40 (0.8)	1527 (29.5)
Unknown	3 (0.05)	3850 (63.8)	0	0
Family history				
Yes	23 (0.4)	45 (0.7)	22 (0.4)	789 (15.2)
No	57 (0.9)	104 (1.7)	89 (1.7)	4256 (82.1)
Unknown	37 (0.6)	5769 (95.6)	0	29 (0.56)
High-risk pre-neoplasia				
Yes	32 (0.5)	574 (9.5)	0	0
No/unknown	85 (1.4)	5344 (88.6)	111 (2.1)	5074 (97.9)
Breast density				
Extreme dense	7 (0.1)	26 (0.4)	0	0
Heterogeneously dense	47 (0.8)	62 (1.03)	0	0
Scattered	14 (0.2)	27 (0.4)	0	0
Almost entirely fat	0	11 (0.2)	0	0
Unknown	49 (0.8)	5792 (96)	111 (2.1)	5074 (97.9)
ER status				
Negative	26 (0.4)	1215 (20.1)	25 (0.5)	944 (18.2)
Positive	75 (1.2)	3818 (63.3)	86 (1.7)	4130 (79.7)
Unknown	16 (0.3)	885 (14.7)	0	0
First BC type				
Pure DCIS	12 (0.2)	1075 (17.8)	3 (0.06)	45 (0.9)
Mixed	74 (1.2)	2450 (40.6)	0	35 (0.7)
Pure invasive	31 (0.5)	2393 (39.7)	108 (2.1)	4994 (96.3)
Age at first birth				
< 30/nulliparous	66 (1.1)	107 (1.8)	10 (0.2)	261 (5.03)
30–40	10(0.2)	29 (0.5)	2 (0.04)	57 (1.1)
40 +	1 (0.02)	2 (0.03)	0	4 (0.1)
Unknown	40 (0.7)	5780 (95.8)	99 (1.9)	4752 (91.7)

The numbers within parentheses indicate percentage of the cohort size

Table 3 shows sensitivity and specificity values at the quintiles of risks predicted by CBCRisk. Also presented are the effective cohort sizes (*n*) that have been used in this calculation after excluding controls with insufficient follow-up periods (as described in Section “Data Sources”). First, we note that, in general, the absolute risks are low. For example, 80% women have less than 2.4 and 1.53% predicted risks of getting CBC in 5 years in the JH and MDA cohorts, respectively. Also, the absolute risks are

higher for JH than MDA; we will elaborate on this point in the Discussion section. We can see from the table that thresholds of 1% or less correspond to sensitivity values between 0.8 and 0.9 depending on the cohort and the number of years in the prediction period.

Figures 1 and 2 present the ROC curves for 3- and 5-year predicted risks for the two cohorts, respectively. Except near the left edge in the MDA cohort, the curves in all cases are above the 45° line, indicating that the model is useful for distinguishing between those who get CBC and those who do not. This is also confirmed by the AUC estimates, which for 3-year risks are 0.65 for JH and 0.61 for MDA. These estimates for 5-year risks are 0.62 and 0.61 for JH and MDA, respectively. The two ROC curves in each figure cross, so the model’s performance in one cohort cannot be said to be uniformly better than the other. However, based on the AUC, the model’s overall discrimination ability is better in the JH cohort than the MDA cohort. A likely reason for this is that information on high-risk pre-neoplasia status and breast density—two risk factors associated with higher risk—is available for some women in JH cohort but is totally missing in MDA cohort. Also, in the JH cohort, the AUC is larger for 3-year risks than for 5-year risks whereas the two are the same in the MDA cohort.

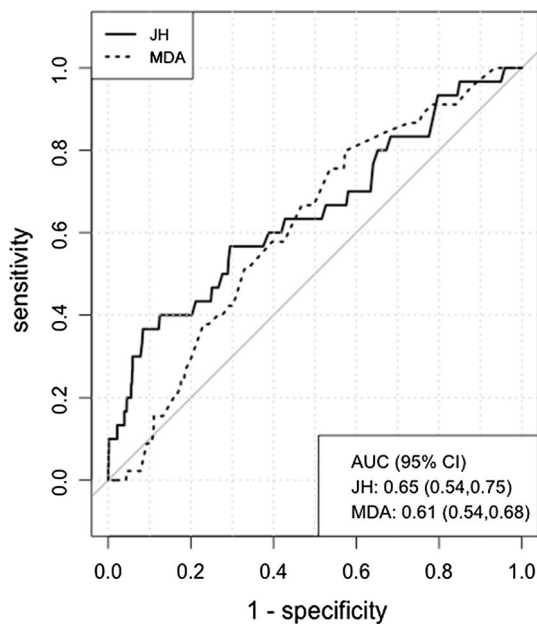
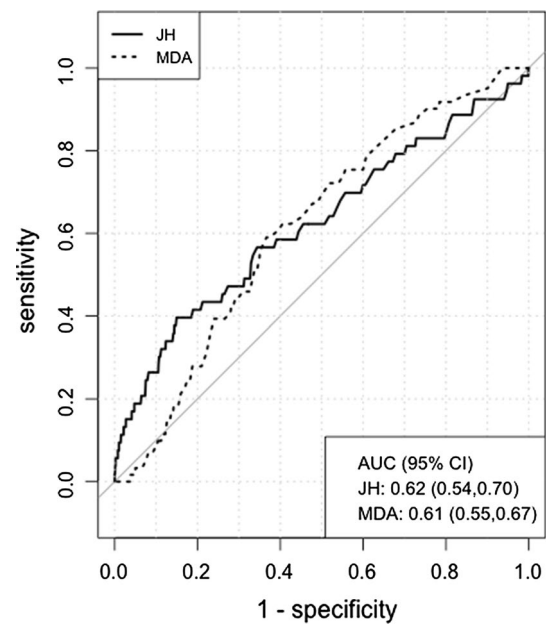
Next, we consider the calibration performance. Tables 4 and 5 present estimates of *E/O* and their 95% CI for 3- and 5-year risks. In addition to the overall *E/O*, the tables show *E/O* for the five risk quintile groups that are formed by the thresholds listed in Table 3, and for different levels of each risk factor used in CBCRisk. For the JH cohort, the overall *E/O* for 3-year risk is 2.02, implying that CBCRisk overpredicts the number of CBC cases by about twofold. The overprediction is persistent across the board in all five risk quintile groups and in categories of age at first diagnosis, ER status, first BC type, and high-risk pre-neoplasia. On the other hand, the risk is underpredicted in most of the categories of family history and age at first birth. Note, however, that only 30 cases of CBC (0.6% of the effective cohort) are observed at the end of the 3-year prediction period, and many subgroups have 5 or fewer cases. The *E/O* in groups that have a very small number of cases may not be stable. The calibration improves for 5-year risk in which case the overall *E/O* is 1.56 and 53 cases (1.3% of the effective cohort) are observed at the end of the 5-year prediction period. Except the categories with a very small number of cases, the pattern of over- and underprediction is similar to that seen for 3-year risks.

For the MDA cohort, the overall *E/O* for 3- and 5-year risks are 0.61 and 0.52, respectively. Thus, in contrast to the JH cohort, the model underpredicts the number of CBC cases in this cohort by 40 and 50% for 3- and 5-year risk predictions, respectively. The underprediction occurs in all risk quintile groups except one as well as in all risk factor

Table 3 Sensitivities and specificities of CBCRisk at thresholds given by the quintiles of predicted 3- and 5-year risks (in %)

Johns Hopkins 3-year risk ($n = 5026$)						
Threshold	0.49	0.90	1.04	1.20	1.47	4.25
Sensitivity	1.00	0.93	0.70	0.60	0.40	0.03
Specificity	0.00	0.18	0.39	0.59	0.80	1.00
MD Anderson 3-year risk ($n = 3547$)						
Threshold	0.45	0.57	0.63	0.76	0.91	4.93
Sensitivity	1.00	0.91	0.84	0.58	0.36	0.00
Specificity	0.00	0.16	0.33	0.60	0.78	1.00
Johns Hopkins 5-year risk ($n = 4142$)						
Threshold	0.81	1.48	1.73	2.01	2.40	7.19
Sensitivity	1.00	0.87	0.72	0.58	0.42	0.02
Specificity	0.00	0.19	0.40	0.60	0.80	1.00
MD Anderson 5-year risk ($n = 2478$)						
Threshold	0.73	0.95	1.08	1.22	1.53	8.19
Sensitivity	1.00	0.92	0.75	0.62	0.28	0.00
Specificity	0.00	0.19	0.40	0.59	0.80	1.00

The first and the last thresholds are the minimum and maximum predicted risks and the thresholds in-between are the four quintiles. The effective cohort sizes are denoted by n

**Fig. 1** ROC curve and AUC for 3-year risks predicted by CBCRisk**Fig. 2** ROC curve and AUC for 5-year risks predicted by CBCRisk

groups. Moreover, the observed numbers of cases at the end of 3- and 5-year prediction periods are 45 and 61 (1.3 and 2.5% of effective cohorts), respectively.

Discussion

In this article, we evaluated the calibration and discrimination performance of CBCRisk for predicting 3- and 5-year absolute risks starting from age at the first BC diagnosis. From the calibration evaluation, we find that CBCRisk

overpredicts the number of CBC cases in the JH cohort by 1.56–2.02-fold, and it underpredicts in the MDA cohort by 40–50%. This difference in performance can be explained to some extent by the difference in the compositions of the two cohorts. The women in the JH cohort tend to be associated with higher risks of CBC compared to those in the MDA cohort, thereby leading to an increased value of E for that cohort. For example, the “pure DCIS” and “mixed invasive/DCIS” categories of first BC type are associated with higher risks of CBC compared to the “pure invasive” category [10]; and the JH cohort is 60% “pure DCIS” or

Table 4 Expected (*E*) and observed (*O*) numbers of CBC cases for 3-year predicted risk together with 95% CI for *E/O*

	Johns Hopkins				MD Anderson			
	<i>n</i>	<i>E</i>	<i>O</i>	<i>E/O</i> (95% CI)	<i>n</i>	<i>E</i>	<i>O</i>	<i>E/O</i> (95% CI)
Overall	5026	60.57	30	2.02 (1.41, 2.89)	3547	27.33	45	0.61 (0.46, 0.82)
Risk quintile groups								
Group 1	1015	7.96	2	3.98 (1.00,15.91)	758	4.12	4	1.03 (0.39,2.74)
Group 2	1010	9.77	7	1.40 (0.67,2.93)	734	4.52	5	0.90 (0.37,2.16)
Group 3	1000	11.23	3	3.74 (1.21,11.61)	641	4.59	10	0.46 (0.25,0.85)
Group 4	1011	13.39	6	2.23 (1.00,4.97)	708	5.87	13	0.44 (0.26,0.77)
Group 5	990	18.22	12	1.52 (0.86,2.67)	706	8.24	13	0.63 (0.37,1.08)
Risk factor groups								
Age at first diagnosis								
< 30	40	0.44	0	–	36	0.25	2	0.12 (0.03,0.48)
30–40	417	4.44	4	1.11 (0.42,2.96)	268	1.82	3	0.61 (0.20,1.89)
40 +	4569	55.69	26	2.14 (1.46,3.15)	3243	25.26	40	0.63 (0.46,0.86)
Anti-estrogen therapy								
No	854	11.31	11	1.03 (0.57,1.86)	911	9.00	18	0.50 (0.32,0.79)
Yes	1086	11.15	17	0.66 (0.41,1.06)	2636	18.32	27	0.68 (0.47,0.99)
Unknown	3086	38.10	2	19.05 (4.76,76.17)	0	0	–	–
Family history								
Yes	68	1.45	9	0.16 (0.08,0.31)	576	6.60	11	0.60 (0.33,1.08)
No	160	2.00	11	0.18 (0.10,0.33)	2947	20.51	34	0.60 (0.43,0.84)
Unknown	4798	57.29	10	5.71 (3.08,10.65)	24	0.22	0	–
High-risk pre-neoplasia								
Yes	531	10.16	7	1.45 (0.69,3.04)				
No/unknown	4495	50.04	23	2.19 (1.46,3.30)				
Breast density								
Extreme dense	33	0.55	3	0.18 (0.06,0.57)				
Heterogeneously dense	108	1.69	10	0.17 (0.09,0.31)				
Scattered	41	0.63	6	0.11 (0.05,0.23)				
Almost entirely fat	11	0.11	0	–				
Unknown	4833	57.59	11	5.24 (2.90,9.45)				
ER status								
Negative	935	12.41	8	1.55 (0.78,3.10)	640	6.39	9	0.71 (0.37,1.36)
Positive	3264	38.7	20	1.94 (1.25,3.00)	2907	20.94	36	0.58 (0.42,0.80)
Unknown	827	9.46	2	4.73 (1.18,18.91)	0	0	–	–
First BC type								
Pure DCIS	1017	15.97	1	15.97 (2.25,113.52)	43	0.61	1	0.61 (0.09,0.80)
Mixed	2126	26.82	22	1.22 (0.80,1.85)	30	0.32	0	–
Pure invasive	1883	17.79	7	2.54 (1.21,5.33)	3474	26.41	44	0.60 (0.45,0.81)
Age at first birth								
< 30/nulliparous	172	2.52	14	0.18 (0.11,0.30)	219	1.65	7	0.24 (0.11,0.05)
30–40	39	0.62	5	0.12 (0.05,0.30)	47	0.45	1	0.45 (0.06,3.19)
40 +	3	0.10	1	0.10 (0.01,0.71)	4	0.17	0	–
Unknown	4812	57.32	10	5.73 (3.08,10.65)	3277	25.06	37	0.68 (0.49,0.94)

If *O* is zero, this CI is undefined and hence is not reported

“mixed invasive/DCIS,” whereas the MDA cohort is 98% “pure invasive.” Likewise, unknown family history of BC is associated with a higher risk compared to no family history and the JH cohort has 96% unknown family history, while the MDA cohort has 83% no family history.

Moreover, unknown status of anti-estrogen therapy is associated with a higher risk compared to having the therapy; and the JH cohort has 63% with unknown status, whereas in the MDA cohort 70% have anti-estrogen therapy.

Table 5 Expected (*E*) and observed (*O*) numbers of CBC cases for 5-year predicted risk together with 95% CI for *E/O*

	Johns Hopkins				MD Anderson			
	<i>n</i>	<i>E</i>	<i>O</i>	<i>E/O</i> (95% CI)	<i>n</i>	<i>E</i>	<i>O</i>	<i>E/O</i> (95% CI)
Overall	4142	82.53	53	1.56 (1.19,2.04)	2478	31.66	61	0.52 (0.40,0.67)
Risk quintile groups								
Group 1	842	10.91	9	1.21 (0.63,2.33)	536	4.73	5	0.97 (0.40,2.33)
Group 2	831	13.31	7	1.90 (0.91,3.99)	512	5.11	10	0.52 (0.28,0.97)
Group 3	840	15.65	6	2.61 (1.17,5.81)	469	5.44	9	0.61 (0.32,1.17)
Group 4	805	17.67	9	1.96 (1.02,3.77)	509	6.89	20	0.35 (0.23,0.54)
Group 5	824	24.99	22	1.14 (0.75,1.73)	500	9.49	17	0.57 (0.35,0.92)
Risk factor groups								
Age at first diagnosis								
< 30	29	0.54	0	–	20	0.25	2	0.13 (0.03,0.52)
30–40	344	6.18	7	0.88 (0.42,1.85)	175	1.99	3	0.67 (0.22,2.08)
40 +	3769	75.81	46	1.65 (1.23,2.20)	23331	30.03	56	0.54 (0.42,0.70)
Anti-estrogen therapy								
No	754	16.46	22	0.75 (0.49,1.14)	654	10.78	25	0.43 (0.29,0.64)
Yes	947	16.08	28	0.58 (0.40,0.83)	1872	21.50	36	0.60 (0.43,0.83)
Unknown	2441	49.99	3	16.66 (5.37,51.67)	0	0	0	–
Family history								
Yes	68	2.38	12	0.20 (0.11,0.35)	433	8.22	14	0.59 (0.35,1.00)
No	160	3.28	24	0.14 (0.09,0.20)	2075	23.80	47	0.51 (0.38,0.68)
Unknown	3914	76.87	17	4.52 (2.81,7.27)	18	0.26	0	–
High-risk pre-neoplasia								
Yes	470	14.69	13	1.13 (0.66,1.95)				
No/unknown	3672	67.84	40	1.70 (1.24,2.31)				
Breast density								
Extreme dense	33	0.91	5	0.18 (0.08,0.44)				
Heterogeneously dense	108	2.77	18	0.15 (0.10,0.24)				
Scattered	41	1.03	10	0.10 (0.06,0.19)				
Almost entirely fat	11	0.18	0	–				
Unknown	3949	77.65	20	3.88 (2.50,6.02)				
ER status								
Negative	735	16.22	15	1.08 (0.65,1.79)	475	7.87	15	0.52 (0.31,0.86)
Positive	2650	51.99	33	1.58 (1.12,2.22)	2051	24.41	46	0.53 (0.40,0.71)
Unknown	757	14.32	5	2.86 (1.19,6.88)	0			
First BC type								
Pure DCIS	917	23.54	4	5.88 (2.21,15.68)	39	0.90	1	0.90 (0.13,6.39)
Mixed	1744	36.07	36	1.00 (0.72,1.39)	22	0.38	0	–
Pure invasive	1481	22.92	13	1.76 (1.02,3.04)	2465	30.99	60	0.52 (0.40,0.67)
Age at first birth								
< 30/nulliparous	172	4.12	29	0.14 (0.10,0.20)	168	2.13	8	0.27 (0.14,0.54)
30–40	39	1.03	6	0.17 (0.08,0.38)	36	0.58	2	0.29 (0.07,1.16)
40 +	3	0.17	1	0.17 (0.02,1.21)	4	0.28	0	–
Unknown	3928	77.21	17	4.54 (2.82,7.31)	2318	28.29	51	0.57 (0.43,0.75)

If *O* is zero, this CI is undefined and hence is not reported

Another reason is related to the requirement that the controls be followed up for at least 3 and 5 years for validation of 3- and 5-year risks, respectively. Obviously, the number of controls satisfying this requirement would drop as the risk prediction period becomes longer. However, the

drop is much steeper for the MDA cohort than for the JH cohort because the follow-up times for controls are generally shorter for MDA. For example, in case of JH, out of 5918 total controls in the cohort, 4996 (84%) have been followed up for at least 3 years and 4089 (69%) for 5 years. In

contrast, for MDA, out of 5074 total controls in the cohort, only 3437 (68%) and 2417 (48%) have been followed up for at least 3 and 5 years, respectively. A smaller number of controls in the MDA cohort leads to a decreased value of E for that cohort. Also, in the MDA cohort, the missing values on high-risk pre-neoplasia and breast density for all women and age at first birth for most women might have contributed to underprediction of number of CBC cases.

Yet another difference between the two cohorts is the prevalence of CBC. It is lower in the JH cohort than that the MDA cohort. The latter is closer to what we see in large population-based databases such as BCSC and SEER. However, the JH cohort spans relatively more recent years (1990–2016) compared to MDA (1983–2014) and BCSC (1995–2009). As the incidence of CBC has declined over the years in general, this might partially explain the lower prevalence, and hence overprediction by CBCRisk, in the JH cohort.

From the discrimination evaluation, we find that the AUC estimates range between 0.61 and 0.65. To put these calibration and discrimination results for CBCRisk into perspective, we may compare them to those reported in the validation studies of the popular BCRAT and related models for predicting a woman's risk of developing first invasive BC. A review of literature shows that the overall E/O ratios reported therein range between 0.5 and 2.5 with an average of 1.3, and the AUC reported therein range between 0.47 and 0.65 with an average of 0.57 [14, 19–25]. Thus, the performance of CBCRisk is in line with those of BCRAT and related models. Given that these models are widely used in practice, it may be reasonable to conclude that CBCRisk should also be considered acceptable for clinical use.

Recently, a set of guidelines from Manchester, UK has been proposed for counseling women about CPM [26]. It includes a simple formula for estimating the *lifetime* risk of CBC. For this, a total risk is first calculated using 0.5% risk per year starting from the age of first diagnosis to age 80 (the formula assumes 80 as average life expectancy) and then the risk is adjusted according to personal factors of anti-endocrine treatment, BRCA mutation carrier status, oophorectomy, and family history. For example, for a woman receiving anti-endocrine treatment, the total risk is reduced by 50% while for positive family history, the risk is doubled. This estimation does not involve formal modeling and does not provide 5- or 10-year risk of CBC, which may be more relevant as most women who get CBC typically get it within 15 years of their first BC diagnosis (in SEER data, we found that over 99% CBC cases occurred within 15 years of first BC diagnosis). Even though CBCRisk can be used to calculate risk up to age 80 or 88 (assuming either one to be lifetime), very few women in the validation cohorts have been followed up for 10 years or more, let alone for their lifetime. Thus,

validating a lifetime risk calculation is problematic with these data. If data with long-term follow-ups as well as relevant risk factors become available, a comparison of Manchester formula and CBCRisk will be worth doing.

A limitation of our study is missing data. A substantial number of women in both cohorts do not have information on several risk factors used in CBCRisk, which may adversely affect the model's ability to predict accurately. Moreover, large proportions of women have not been followed up for more than 5 years. There is, in general, a dearth of prospectively collected data on BC patients that have all relevant risk factors recorded on most women. For example, breast density is typically unavailable in earlier data and age at first birth may not have been collected on women who have been already diagnosed with breast cancer (as opposed to unaffected women for whom age at first birth is used for risk prediction of first BC and hence is collected). Despite the limitations, our study shows that CBCRisk is useful for prediction of risk of CBC for woman diagnosed with unilateral breast cancer. Thus, we believe that CBCRisk can be used confidently in clinical settings to aid physicians in quantifying the risk of CBC, and thereby help BC patients weigh the pros and cons of various treatment and prophylactic options in a more informed manner. An effective counseling aided with a personalized CBC risk estimate may potentially help alleviate the rate of medically unnecessary CPMs.

Funding This work was funded by the National Cancer Institute at the National Institutes of Health (Grant Number R21CA186086).

References

1. Tuttle TM, Habermann EB, Grund EH, Morris TJ, Virnig BA (2007) Increasing use of contralateral prophylactic mastectomy for breast cancer patients: a trend toward more aggressive surgical treatment. *J Clin Oncol* 25(33):5203–5209
2. Tuttle TM, Jarosek S, Habermann EB, Arrington A, Abraham A, Morris TJ, Virnig BA (2009) Increasing rates of contralateral prophylactic mastectomy among patients with ductal carcinoma in situ. *J Clin Oncol* 27(9):1362–1367
3. King TA, Sakr R, Patil S, Gurevich I, Stempel M, Sampson M, Morrow M (2011) Clinical management factors contribute to the decision for contralateral prophylactic mastectomy. *J Clin Oncol* 29(16):2158–2164
4. Wong SM, Freedman RA, Sagara Y, Aydogan F, Barry WT, Golshan M (2017) Growing use of contralateral prophylactic mastectomy despite no improvement in long-term survival for invasive breast cancer. *Ann Surg* 265(3):581–589
5. Nichols HB, Berrington de Gonzalez A, Lacey JV, Rosenberg PS, Anderson WF (2011) Declining incidence of contralateral breast cancer in the United States from 1975 to 2006. *J Clin Oncol* 29(12):1564–1569
6. Murphy JA, Milner TD, O'Donoghue JM (2013) Contralateral risk-reducing mastectomy in sporadic breast cancer. *Lancet Oncol* 14(7):e262–269

7. Abbott A, Rueth N, Pappas-Varco S, Kuntz K, Kerr E, Tuttle T (2011) Perceptions of contralateral breast cancer: an overestimation of risk. *Ann Surg Oncol* 18(11):3129–3136
8. Metcalfe KA, Narod SA (2002) Breast cancer risk perception among women who have undergone prophylactic bilateral mastectomy. *J Natl Cancer Inst* 94(20):1564–1569
9. Khan SA (2011) Contralateral prophylactic mastectomy: what do we know and what do our patients know? *J Clin Oncol* 29(16):2132–2135
10. Chowdhury M, Euhus D, Onega T, Biswas S, Choudhary PK (2017) A model for individualized risk prediction of contralateral breast cancer. *Breast Cancer Res Treat* 161:153–160. <https://doi.org/10.1007/s10549-016-4039-x>
11. Breast Cancer Surveillance Consortium. last accessed on April 8, 2016. Data retrieved BCSC, <https://breastscreening.cancer.gov/>
12. Surveillance, Epidemiology, and End Results Program. Research data (1973–2010). released April 2013, based on the November 2012. <https://seer.cancer.gov>
13. CBCRisk. last accessed on February 15, 2018. <http://www.utdallas.edu/swati.biswas/>
14. Spiegelman D, Colditz GA, Hunter D, Hertzmark E (1994) Validation of the Gail et al. model for predicting individual breast cancer risk. *J Natl Cancer Inst* 86(8):600–607
15. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3):837–845
16. Carpenter J, Bithell J (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 19(9):1141–1164
17. R Development Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016. <http://www.R-project.org>
18. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Miller M (2011) pROC: an open-source package for R and S+ to analyze and compare roc curves. *BMC Bioinform* 12:77
19. Bondy ML, Lustbader ED, Halabi S, Ross E, Vogel VG (1994) Validation of a breast cancer risk assessment model in women with a positive family history. *J Natl Cancer Inst* 86(8):620–625
20. Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, Wieand HS (1999) Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst* 91(18):1541–1548
21. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA (2001) Model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 93(5):358–366
22. Chay WY, Ong WS, Tan PH, Leo NQJ, Ho GH, Wong CS, Chia KS et al (2012) Validation of the Gail model for predicting individual breast cancer risk in a prospective nationwide study of 28,104 Singapore women. *Breast Cancer Res* 14(1):R19
23. Min JW, Chang MC, Lee HK, Hur MH, Noh DY, Yoon JH, Jung Y, Yang JH (2014) Validation of risk assessment models for predicting the incidence of breast cancer in Korean women. *J Breast Cancer* 17(3):226–235
24. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Srinakaran J, Woodtichartpreecha P, Hirunpat S, Wongwaisayawan S, Lertsithichai P, Kasamesup V, Thakkinstian A (2014) Development and validation of a breast cancer risk prediction model for Thai women: a cross-sectional study. *Asian Pac J Cancer Prev* 15(16):6811–6817
25. Pankratz VS, Degnim AC, Frank RD, Frost MH, Visscher DW, Vierkant RA, Hieken TJ, Ghosh K, Tarabishy Y, Vachon CM et al (2015) Model for individualized prediction of breast cancer risk after a benign breast biopsy. *J Clin Oncol* 33(8):923–929
26. Basu NN, Ross GL, Evans DG, Barr L (2015) The Manchester guidelines for contralateral risk-reducing mastectomy. *World J Surg Oncol* 13:237. <https://doi.org/10.1186/s12957-015-0638-y>