PredictCBC-2.0: a contralateral breast cancer risk prediction model developed and validated in ~ 200,000 patients

Daniele Giardiello1,2,3, Maartje J. Hooning4, Michael Hauptmann5, Renske Keeman1, B. A. M. Heemskerk-Gerritsen4, Heiko Becher6, Carl Blomqvist7,8, Stig E. Bojesen9,10,11, Manjeet K. Bolla12, Nicola J. Camp13, Kamila Czene14, Peter Devilee15,16, Diana M. Eccles17, Peter A. Fusching18,19, Jonine D. Figueroa20,21,22, Henrik Flyger23, Montserrat García-Closas22, Christopher A. Haiman24, Ute Hamann25, John L. Hopper26, Anna Jakubowska27,28, Floor E. Leeuwen29, Annika Lindblom30,31, Jan Lubiński27, Sara Margolin32,33, Maria Elena Martinez34,35, Heli Nevanlinna36, Ines Nevelsteen37, Saskia Pelders4, Paul D. P. Pharoah12,38, Sabine Siesling39,40, Melissa C. Southey41,42,43, Annemieke H. van der Hout44, Liselotte P. van Hest45, Jenny Chang-Claude46,47, Per Hall14,32, Douglas F. Easton12,38, Ewout W. Steyerberg2,48 and Marjanka K. Schmidt1,29*

Abstract

Background: Prediction of contralateral breast cancer (CBC) risk is challenging due to moderate performances of the known risk factors. We aimed to improve our previous risk prediction model (PredictCBC) by updated follow-up and including additional risk factors.

Methods: We included data from 207,510 invasive breast cancer patients participating in 23 studies. In total, 8225 CBC events occurred over a median follow-up of 10.2 years. In addition to the previously included risk factors, PredictCBC-2.0 included CHEK2 c.1100delC, a 313 variant polygenic risk score (PRS-313), body mass index (BMI), and parity. Fine and Gray regression was used to fit the model. Calibration and a time-dependent area under the curve (AUC) at 5 and 10 years were assessed to determine the performance of the models. Decision curve analysis was performed to evaluate the net benefit of PredictCBC-2.0 and previous PredictCBC models.

Results: The discrimination of PredictCBC-2.0 at 10 years was higher than PredictCBC with an AUC of 0.65 (95% prediction intervals (PI) 0.56–0.74) versus 0.63 (95%PI 0.54–0.71). PredictCBC-2.0 was well calibrated with an observed/expected ratio at 10 years of 0.92 (95%PI 0.34–2.54). Decision curve analysis for contralateral preventive mastectomy (CPM) showed the potential clinical utility of PredictCBC-2.0 between thresholds of 4 and 12% 10-year CBC risk for BRCA1/2 mutation carriers and non-carriers.
**Introduction**

Contralateral breast cancer (CBC) is the most common second primary cancer among women diagnosed with first primary invasive breast cancer (BC) [1]. CBC accounts for approximately 40–50% of all new secondary cancers in women with first primary invasive BC and has a potentially less favorable prognosis [2–6]. Worries regarding CBC risk have increased the demand for contralateral preventive mastectomy (CPM) [7, 8]. However, the impact of CPM on survival is uncertain, especially in women with a low risk to develop a CBC [9–13]. Thus, improved CBC risk prediction is important in order to inform decision-making on surveillance and preventive strategies. Currently, the most important factor for decision-making on CPM is the BRCA1/2 mutation status [14].

We previously developed and cross-validated two models using data from 132,756 invasive BC patients with a median follow-up of 8.8 years including 4672 CBC events [15]. One model (PredictCBC-1A) was developed including information about BRCA1/2 mutation status and another model (PredictCBC-1B) for the general breast cancer population of genetically untested women. Two other specific CBC prediction tools are currently available in the literature: the Manchester formula (part of the Manchester guidelines for CPM) and CBCrisk [15–18].

In addition to BRCA1/2 mutations, other genetic risk factors for breast cancer are also associated with CBC risk. In particular, there is substantial evidence that the CHEK2 c.1100delC variant increases the risk of developing CBC [19, 20]. In addition, polygenic risk scores (PRS) of common variants, developed for association with first breast cancer, have been shown to predict CBC in the general BC population and in BRCA1/2 mutation carriers [21–24], particularly the extensively validated 313 SNP PRS [25]. With regard to the lifestyle and reproductive factors, there is evidence that body mass index (BMI) and parity at or around the time of the first primary invasive BC diagnosis are associated with CBC risk [26].

Our aim was to refit PredictCBC models incorporating these additional risk factors. We utilized the same dataset but with updated follow-up and added additional studies, especially one large study of BRCA1 and BRCA2 mutation carriers. We evaluated the potential improvement in prediction performance and utility for clinical decision-making of the updated models for both BRCA1/2 carriers as the general (non-tested) breast cancer population (PredictCBC-2.0).

**Material and methods**

**Study population and available data**

We used the data from the same five main sources previously used for PredictCBC models to develop the PredictCBC-2.0 models including updated follow-up information, additional patients, and invasive or in situ CBC events [15]. Two studies were additionally included from the Breast Cancer Association Consortium (BCAC) compared to the version of the BCAC data used to develop PredictCBC-1A and PredictCBC-1B models. Most of the studies were either population- or hospital-based series; and most women were of European descent (Additional file 1: Data and patient selection and Additional file 2: Table S1 and Additional file 1: Table S2, available online). We also additionally included patients selected from the Hereditary Breast and Ovarian cancer study in the Netherlands (HEBON) [27], a nationwide study based on clinical genetic centers. The eligibility criteria were the same as previously: briefly, we included female patients with invasive first primary BC with no sign of distant metastases at diagnosis or prior history of any cancer (except for non-melanoma skin cancer) [15]. We included women diagnosed after 1990 so that diagnostic and treatment procedures were close to modern practice while follow-up was sufficient to study CBC incidence. In total, 207,510 women with first primary invasive BC from 23 studies were included. All studies were approved by the appropriate ethics and scientific review boards. All women provided written informed consent; or, for some Dutch cohorts as applicable, the secondary use of clinical data was in accordance with Dutch legislation and codes of conduct [28, 29]. Information about the sample size for every data source and the total sample size after eligibility criteria are provided in Table 1. The choice of additional predictors in the analyses was based on evidence from the literature and the availability of predictors in our data sources. In particular, evidence from the literature suggests that CHEK2 c.1100delC and 313 SNP PRS increased the risk of developing CBC...
In addition, a systematic review of lifestyle and reproductive factors suggested that BMI and parity at or around the time of the first primary invasive BC diagnosis are associated with CBC risk [26]. Details about sample size per study and about the factors included in the analyses, follow-up per dataset, and study design are in Additional file 2: Table S1 and Additional file 3: Table S3, available online.

**Statistical analyses**

**Primary endpoint and follow-up**

The primary endpoint in the analyses was the incidence of invasive or in situ metachronous CBC. Follow-up started 3 months after invasive first primary BC diagnosis, to exclude synchronous CBCs, and ended at the date of CBC, distant metastasis (but not a loco-regional relapse), CPM, or last date of follow-up (due to death, loss to follow-up, or end of study), whichever occurred first. For 36,553 (17.6%) women, from BCAC and HEBON, recruitment or blood sampling for DNA testing occurred more than 3 months after diagnosis of the first primary BC. For women with the first primary invasive BC, follow-up started at recruitment or at the date of blood draw or at DNA test result (left truncation). Patients who underwent CPM during the follow-up were censored because of negligible CBC risk after a CPM [30]. Missing data were multiply imputed by chained equations (MICE) to avoid loss of information due to case-wise deletion [31–33] (Additional file 1: Multiple imputation of missing values, available online).

**Model development and validation**

We used multivariable Fine and Gray regression models to account for death and distant metastases as competing events [34]. Analyses were stratified by a study to allow baseline hazard (sub)distributions to differ across studies. The assumption of proportional subdistribution hazards was graphically checked using Schoenfeld residuals [35]. The resulting subdistribution hazard ratios (sHRs) and corresponding 95% confidence intervals (CI) were pooled from 5 imputed datasets using Rubin’s rules [33].

We re-estimated the coefficients of PredictCBC-1A and PredictCBC-1B, and we re-fitted the PredictCBC models using the extended dataset with updated follow-up time. PredictCBC-1A, developed including information about BRCA1/2 germline mutation after CBC or preventive mastectomy, was extended by including CHEK2 c.1110delC status, PRS-313, self-reported BMI, and self-reported parity (hereafter: PredictCBC-2.0A) [15]. CHEK2 c.1110delC and PRS-313 were derived from the BCAC database, as published previously [25, 36, 37]. We extended PredictCBC-1B, developed for genetically untested women, incorporating...
self-reported BMI and parity (hereafter: PredictCBC-2.0B). Potential nonlinear relations between continuous predictors and CBC risk were investigated using restricted cubic splines with three knots.

The validity of the model was investigated by leave-one-study-out cross-validation [38]. In each validation cycle, all studies were analyzed except one, in which the validity of the model was evaluated. Since some BCAC studies had insufficient CBC events required for reliable validation, we used the geographic area as a unit for splitting [38–40]. Nineteen out of 23 studies were combined in 4 geographic areas (Additional file 1: Table S2, available online). A total of 8 units of splitting including 4 geographic areas and 4 studies were used to cross-validate the models.

The performance of the PredictCBC-2.0 was assessed by discrimination, i.e., the ability to differentiate between patients diagnosed with CBC and those who were not, and by calibration, which measures the agreement between the actual (observed) risk and CBC risk estimated by the prediction models (predicted). Discrimination was quantified by time-dependent areas under the ROC curve (AUCs) based on Inverse Censoring Probability Weighting at 5 and 10 years [41]. The AUCs were estimated using the prognostic index which is a combination of the estimated coefficients (betas) of PredictCBC models multiplied by the corresponding individual characteristics (i.e., predictors) included in the models. Values of AUCs close to 1 indicate good discrimination, while values close to 0.5 indicate poor discrimination. Calibration was assessed by the observed-to-expected (O/E) ratio and calibration plots at 5 and 10 years [42, 43]. An O/E ratio lower or higher than 1 indicates that average predictions are too high or low, respectively.

To consider heterogeneity among studies, a random-effect meta-analysis was performed to provide summaries of discrimination and calibration performance. The 95% prediction intervals (PI) indicate the likely performance of the model in a new dataset. The summary performances of PredictCBC-2.0 and 1.0 models were compared to evaluate whether adding the new predictors improved the performance of CBC risk prediction. We developed and validated the risk prediction model following the Transparent Reporting of a Multivariable Prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement [44]. Analyses were done in SAS (SAS Institute Inc., Cary, NC, USA) and R (version 3.6.1).

Clinical utility
The clinical utility of the prediction models was evaluated using decision curve analysis (DCA) [45, 46]. A key metric DCA is the net benefit, which is the number of true-positive classifications (in this example: the number of CPMs in patients who would have developed a CBC) minus the weighted number of false-positive classifications (in this example: the number of unnecessary CPMs in patients who would not have developed a CBC). The false positives are weighted by a factor related to the relative harm of a missed CBC versus an unnecessary CPM. The weighting is derived from the threshold probability to develop a CBC using a fixed time horizon (e.g., CBC risk at 5 or 10 years) [47]. For example, a threshold of 10% implies that CPM in 10 patients, of whom one would develop CBC if untreated, is acceptable (thus performing 9 unnecessary CPMs). The net benefit of a prediction model is traditionally compared with the strategies of treat all or treat none. Since the use of CPM is generally only considered among BRCA1/2 mutation carriers, the decision curve analysis was reported among BRCA1/2 mutation carriers and non-carriers separately [48]. Among patients not tested for BRCA1/2 germline mutations, we assumed that the decision for CPM is based on family history of breast cancer. The net benefits of PredictCBC-2.0A and PredictCBC-2.0B were compared with the net benefit of PredictCBC-1A and 1B, respectively, to assess the potential improvement in the clinical utility of the updated models.

Results
A total of 207,510 women with invasive first primary BC diagnosed between 1990 and 2017, with 8225 CBC events (6828 invasive, 1397 in situ), from 23 studies, were used for CBC risk prediction modeling (Additional file 2: Table S1, available online). Median follow-up time was 10.2 years, and CBC cumulative incidences at 5 and 10 years were 2.2% and 4.1%, respectively. Details of the studies and patient, tumor, and treatment characteristics are provided in Additional file 3: Table S3 (available online). The multivariable models with estimates for all included factors are given in Table 2.

Most of the factors were independently associated with CBC risk, including the new factors incorporated in the PredictCBC-2.0 models, i.e., s BMI, parity, CHEK2 c.1110delC, and PRS-313. There was no evidence against log-linear relationships between BMI, parity and PRS-313 and CBC risk. Nonlinearity between age at first BC diagnosis and CBC risk was accounted for with a linear spline at age 60 years. The formulae of the PredictCBC models are provided in Additional file 1: Formula to estimate the contralateral breast cancer risk using PredictCBC-2.0A and PredictCBC-2.0B (available online). To calculate the predicted CBC cumulative incidence, we used the event-free baseline probability of the Netherlands Cancer Registry (NCR), as previously [15].
The AUCs at 5 and 10 years of PredictCBC-2.0A were higher than of PredictCBC-1A at 5 years: 0.66, 95% prediction interval (PI) 0.55–0.76 versus 0.62 (95%PI 0.51–0.74); and at 10 years: 0.65 (95%PI 0.56–0.74) versus 0.63 (95%PI 0.54–0.71) (Figs. 1 and 2, Table 3). The AUCs for PredictCBC-2.0B and PredictCBC-1B were both 0.59 (95%PI: PredictCBC-2.0B: 0.51–0.68; PredictCBC-1B:0.49–0.69) at 5 years and both 0.58 (95%PI 0.51–0.65) at 10 years (Figs. 1 and 2, Table 3).

The O/E ratio at 5 and 10 years across all versions of PredictCBC models ranged between 0.90 and 0.92 with similar 95%PIs (Figs. 1 and 2, Table 3). Calibration plots of PredictCBC-2.0 models are provided in Additional file 1: Figs, S1–S4 (available online).

The decision curves showed the net benefit for a range of harm–benefit thresholds at 10-year CBC risk (Fig. 3). We evaluated the potential clinical utility of PredictCBC-2A versus PredictCBC-1.0A for decision thresholds between 4 and 12% for the 10-year CBC risk among BRCA1/2 mutation carriers and non-carriers (Figs. 3 and 4, Table 4). For example, if consensus guidelines would indicate the acceptability of 1 in 10 patients for whom a CPM is recommended developing CBC, a risk threshold of 10% may be used to define high- and low-risk BRCA1/2 mutation carriers based on the absolute 10-year CBC risk prediction estimated by the models. Compared with a strategy recommending CPM to all BRCA1/2 mutation carriers, PredictCBC-1A avoids 76.9 net CPMs per 1000 patients (Table 4). An additional 50.0 CPMs may be avoided using PredictCBC-2.0A compared to PredictCBC-1A. In contrast, almost no non-BRCA1/2 mutation carriers had predictions above the 10% threshold (general BC population, Table 4); three necessary CPMs per 1000 patients would be indicated using PredictCBC-2.0A. Analyses for PredictCBC-1B and PredictCBC-2.0B at 10 years suggested a potential

| Factor (reference) | PredictCBC-2.0A | PredictCBC-2.0B |
|--------------------|-----------------|-----------------|
| Age at PBC, years (75th vs. 25th quartile: 66 vs. 48) | 0.87a (0.83–0.90) | 0.82a (0.78–0.85) |
| Body mass index, kg/m² (75th vs. 25th quartile: 28.4 vs. 22.7) | 1.06 (1.03–1.09) | 1.06 (1.03–1.09) |
| Parity (75th vs. 25th quartile: 3 vs. 1) | 0.85 (0.82–0.88) | 0.86 (0.83–0.90) |
| First-degree family history of BC (yes) | 1.17 (1.12–1.23) | 1.35 (1.29–1.42) |
| BRCA mutation | | |
| BRCA1 versus non-carrier | 4.79 (4.43–5.17) | – |
| BRCA2 versus non-carrier | 3.09 (2.72–4.25) | – |
| PRS313 b (75th vs. 25th quartile: -0.49 vs. 0.32) | 1.35 (1.31–1.39) | – |
| CHEK2 c.1100delC mutation (present) | 2.75 (2.85–3.34) | – |
| Nodal status of PBC (positive) | 0.99 (0.93–1.05) | 0.99 (0.93–1.04) |
| Tumor size category of PBC, cm | | |
| (2.5) versus ≤ 2 | 0.99 (0.94–1.05) | 1.01 (0.96–1.07) |
| > 5 versus ≤ 2 | 1.23 (1.10–1.36) | 1.22 (1.09–1.36) |
| Morphology of PBC (lobular including mixed) | 1.19 (1.12–1.27) | 1.17 (1.10–1.24) |
| Grade of PBC | | |
| Moderately differentiated vs. well differentiated (II vs. I) | 0.93 (0.88–0.99) | 0.98 (0.93–1.04) |
| Poorly differentiated vs. well differentiated (III vs. I) | 0.85 (0.79–0.91) | 0.95 (0.88–1.01) |
| Chemotherapy (yes) | 0.75 (0.70–0.80) | 0.75 (0.70–0.80) |
| Radiotherapy to the breast (yes) | 0.93 (0.89–0.98) | 0.95 (0.90–0.99) |
| ER with endocrine therapy | | |
| Negative/no versus positive/yes | 1.53 (1.43–1.65) | 1.78 (1.67–1.90) |
| Positive/no versus positive/yes | 1.95 (1.83–2.07) | 1.94 (1.82–2.06) |
| HER2 with trastuzumab therapy | | |
| Negative/no versus positive/yes | 1.22 (1.09–1.38) | 1.30 (1.15–1.46) |
| Positive/no versus positive/yes | 1.12 (0.97–1.28) | 1.14 (1.00–1.31) |

vs. versus, sHR subdistributional hazard ratio, CI confidence interval, PRS polygenic risk score, BC breast cancer, PBC first primary breast cancer, ER estrogen receptor, HER2 human epidermal growth factor 2

a age was parameterized as a linear spline with one interior knot at 60 years. For representation purposes, we here provide the sHR for the 75th versus the 25th percentile

b PRS standardized by the same standard deviation (SD) used by Mavaddat et al. (SD = 0.61)[25]
clinical utility between 4 and 6% 10-year CBC risk for patients with and without family history (Table 4 and Figs. 3 and 4). No remarkable improvement in net benefit was detected using PredictCBC-2.0B compared to PredictCBC-1B in decision-making regarding CPM (Table 4 and Fig. 3). Decision curves for CBC risk using PredictCBC and PredictCBC-2.0 at 5 years and the corresponding clinical utility showed similar patterns (Additional file 1: Figs. S5-S6 and Table S4, available online).

Discussion

We evaluated the potential improvement in CBC risk prediction by adding established genetic (CHEK2 c.1100delC and PRS-313) and lifestyle (BMI and parity) factors to the previous PredictCBC models and used additional follow-up information and new studies to provide more reliable estimates.

The current clinical recommendations of CPM are mostly based on the presence of a pathogenic mutation in BRCA1/2 [49, 50]. This seems a reasonable approach according to CBC risk predictions based on the PredictCBC models: few non-BRCA1/2 carriers exceed a 10% 10-year risk threshold. However, approximately 40% of BRCA1/2 mutation carriers do not reach this threshold either, suggesting that a significant proportion of BRCA1/2 carriers might be spared CPM. Additional genetic information beyond BRCA1/2 germline mutation such as the presence of the CHEK2 c.1110delC variant and PRS-313 might improve decision-making.

Currently available CBC models, such as CBCrisk and the Manchester formula, show only moderate discrimination [51]. In addition, the Manchester formula has been shown to systematically overestimate CBC risk [51]. The BOADICEA model, a well-known risk prediction tool to estimate the risk of developing the first primary BC, also allows the calculation of CBC risk [52–55]. Although BOADICEA includes rare pathogenic variants in moderate- and high-risk BC susceptibility genes (i.e., BRCA1, BRCA2, PALB2, ATM and CHEK2, BARD1, RAD51C, RAD51D), and PRS-313, it does not incorporate information on the systemic treatment of the primary BC, which are important predictors of CBC risk [56].
A model for the prediction of recurrence, the INFLU-ENCE nomogram, was developed to estimate 5-year recurrence risk as well as conditional annual risks of developing a local or regional recurrence based on first BC and treatment characteristics [57]. A more recent version (INFLUENCE 2.0) also provides 5-year individualized predictions for secondary primary breast cancer based on cases older than 50 years at first cancer diagnosis from the NCR nationwide cohort irrespective of their genetic status or testing status using random survival forests [58]. The model provided moderate discrimination (AUC at 5 years: 0.67; 95%CI

Table 3  Summary of prediction performance of PredictCBC-1A, PredictCBC-1B, PredictCBC-2.0A, and PredictCBC-2.0B with the corresponding 95% prediction intervals (PI) based on a leave-one-study-out cross-validation procedure

| CBC risk prediction model | Performance measure | Discrimination | Calibration |
|---------------------------|---------------------|----------------|-------------|
|                           |                     | AUC (95% PI)   | O/E ratio (95% PI) |
|                           |                     | 5-year         | 10-year      | 5-year       | 10-year       |
| PredictCBC-1A             | 0.62 (0.51–0.74)    | 0.63 (0.54–0.71) | 0.90 (0.36–2.24) | 0.91 (0.34–2.48) |
| PredictCBC-2.0A           | 0.66 (0.55–0.76)    | 0.65 (0.56–0.74) | 0.91 (0.35–2.34) | 0.92 (0.34–2.54) |
| PredictCBC-1B             | 0.59 (0.49–0.69)    | 0.58 (0.51–0.65) | 0.91 (0.32–2.55) | 0.92 (0.30–2.80) |
| PredictCBC-2.0B           | 0.59 (0.51–0.68)    | 0.58 (0.51–0.65) | 0.91 (0.31–2.63) | 0.92 (0.30–2.87) |

AUC area under the curve, CBC contralateral breast cancer, PI prediction interval, O/E observed/expected
using internal validation. In our comparable population- and hospital-based Dutch series, EMC and NCR, the AUCs at 5 years of PredictCBC-1A were 0.69 (95%CI 0.64–0.73) and 0.66 (95%CI 0.65–0.67), and of PredictCBC-2.0A 0.71 (95%CI 0.66–0.75) and 0.68 (95%CI 0.66–0.69), respectively. Moreover, INFLUENCE 2.0 is only relevant to the general population, while PredictCBC can also be used in the clinical genetic setting. Notably, we demonstrated that decision-making about preventive strategies in clinical practice is unlikely to improve without genetic information.

Our work has some limitations: firstly, some women included in the Dutch studies (providing specific information on family history, BRCA mutation or CPM) were also present in our selection of the NCR population, as described previously [15]. Privacy and coding issues prevented linkage at the individual patient level, but based on the hospitals from which the studies were recruited, and the age and period criteria used, we calculated a maximum potential overlap of 9%. Secondly, important predictors such as family history, BRCA1/2 and CHEK2 c.1110delC status, andPRS-313, were only available in a subset of the women, although the multiple imputation

---

**Fig. 3** Decision curve analysis at 10 years for the contralateral breast cancer risk (CBC) models (PredictCBC-1.0 and PredictCBC-2.0 models) including BRCA mutation information. **A** The decision curve to determine the net benefit of the estimated 10-year predicted CBC cumulative incidence for patients without a BRCA1/2 gene mutation using PredictCBC-1A (dotted black line) and PredictCBC-2.0A (dashed black line) compared to not treating any patients with contralateral preventive mastectomy (CPM) (black solid line). **B** The decision curve to determine the net benefit of the estimated 10-year predicted CBC cumulative incidence for BRCA1/2 mutation carriers using PredictCBC-1A (dotted black line), PredictCBC-2.0A (dashed black line) versus treating (or at least counseling) all patients (gray solid line). **C** The decision curve to determine the net benefit of the estimated 10-year predicted CBC cumulative incidence for patients without (first degree) family history using PredictCBC-1B (dotted black line), PredictCBC-2.0B (dashed black line) compared to not treating any patients with CPM (black solid line). **D** The decision curve to determine the net benefit of the estimated 10-year predicted CBC cumulative incidence for patients with (first degree) family history using PredictCBC-1B (dotted black line), PredictCBC-2.0B (dashed black line) versus treating (or at least counseling) all patients (gray solid line). The y-axis measures net benefit, which is calculated by summing the benefits (true positives, i.e., patients with a CBC who needed a CPM) and subtracting the harms (false positives, i.e., patients with CPM who do not need it). The latter are weighted by a factor related to the relative harm of a non-prevented CBC versus an unnecessary CPM. The factor is derived from the threshold probability to develop a CBC at 10 years at which a patient would opt for CPM (e.g., 10%). The x-axis represents the threshold probability. Using a threshold probability of 10% implicitly means that CPM in 10 patients of whom one would develop a CBC if untreated is acceptable (9 unnecessary CPMs, harm-to-benefit ratio 1:9).
approach should lead to consistent estimates [59–61]. Detailed information about family history of breast cancer would have been useful to improve CBC risk prediction, especially among patients with a mutation in \textit{BRCA1/2} or \textit{CHEK2}. Nonetheless, we considerably increased the number of patients with \textit{BRCA1/2} mutation status and family history information compared to our previous publication (40,343 vs. 7704 and 53,399 vs. 3,548).

\textbf{Fig. 4} Density distribution of 10-year predicted contralateral breast cancer using PredictCBC version 2 models. \textbf{A} Density distribution of 10-year predicted contralateral breast cancer absolute risk using PredictCBC-2.0A within non-carriers (area with black solid lines) and \textit{BRCA1/2} mutation carriers (area with black dashed lines). \textbf{B} Density distribution of 10-year predicted contralateral breast cancer absolute risk using PredictCBC-2.0B within patients without (first degree) family history (area with black solid lines) and patients with (first degree) family history (area with black dashed lines).
vs. 30,541 patients with available BRCA mutation status and family history information, respectively), and added CHEK2 c.1110delC, which is a founder mutation present in approximately 0.5–1.6% of individuals of Northern and Eastern European descent and explains the large majority of carriers of CHEK2 protein truncating variants in these populations [19, 62]. Further validation will be required to investigate how well PredictCBC models predict risk in other populations. In particular, the model was developed in patients of European ancestry and further evaluation and adaptation will be needed to extend PredictCBC models to non-European populations, including Asia [63, 64]. Future research might also include comparisons of machine learning (ML) methods with classical statistical regression models [65, 66].

The prediction models may be further improved by including additional risk factors. In particular, rare mutations in other breast cancer susceptibility genes, such as ATM and PALB2, are also likely to be associated with an increased risk of CBC [22, 67, 68]. The discrimination provided by the PRS will also improve as more SNPs are added [69, 70]. Prediction performance might also be improved by adding breast density and other risk factors (e.g., additional lifestyle and reproductive factors such as alcohol use, age at primiparity, age at menopause) modeled dynamically in a time-dependent fashion [71]. Finally, we wish to emphasize that adequate presentation (e.g., with online tools) of the risk estimates is crucial for effective communication about CBC risk during doctor–patient consultations [72, 73].

**Conclusions**

In conclusion, we present an updated version of a previously proposed contralateral breast cancer risk model (PredictCBC) including additional information on breast cancer genetic variants beyond BRCA1/2, lifestyle and reproductive factors. PredictCBC-2.0, available online at [74], is based on longer follow-up from a wide range of new European-descent population and

| Proportion | PredictCBC-1A and PredictCBC-2.0A | PredictCBC-1B and PredictCBC-2.0B |
|-----------|----------------------------------|----------------------------------|
| **Probability threshold** | **Unnecessary CPM needed to detect one necessary CPM** | **Unnecessary CPMs needed to detect one necessary CPM** |
| **BRCA1/2 mutation carriers** | **Net benefit versus treat all patients with CPM (per 1000)** | **Avoided unnecessary CPMs per 1000 patients using PredictCBC-1A** | **Additional avoided unnecessary CPMs per 1000 patients using PredictCBC-2.0A** | **Net benefit versus treat all patients with CPM (per 1000)** | **Avoided unnecessary CPMs per 1000 patients using PredictCBC-1B** | **Additional avoided unnecessary CPMs per 1000 patients using PredictCBC-2.0B** |
| **Non-carriers** | **Net benefit versus treat all patients with CPM (per 1000)** | **Avoided unnecessary CPMs per 1000 patients using PredictCBC-1B** | **Additional avoided unnecessary CPMs per 1000 patients using PredictCBC-2.0B** | **Net benefit versus treat all patients with CPM (per 1000)** | **Avoided unnecessary CPMs per 1000 patients using PredictCBC-1B** | **Additional avoided unnecessary CPMs per 1000 patients using PredictCBC-2.0B** |
| 4 | 24 | 0.1 | 0.3 | 1.9 | 4.8 | 115.7 | 15.3 |
| 6 | 15.7 | No benefit | 0.0 | 20.0 | 0.6 | 9.3 | 22.9 |
| 8 | 11.5 | 3.5 | 40.6 | 52.0 | No benefit | 0.0 | 9.0 |
| 10 | 9.0 | 8.5 | 76.9 | 50.2 | No benefit | 0.0 | 3.4 |
| 12 | 7.3 | 22.4 | 164.0 | 15.0 | No benefit | 0.0 | 1.1 |

For PredictCBC versions 1A and 2.0A, at the same probability threshold, the net benefit is exemplified in BRCA1/2 mutation carriers (for avoiding unnecessary CPM) and non-carriers (performing necessary CPM). For PredictCBC versions 1B and 2.0B, at the same probability threshold, the net benefit is exemplified in patients with family history (for avoiding unnecessary CPM) and patients without family history (performing necessary CPM).

CPM contralateral preventive mastectomy

*The number of unnecessary contralateral mastectomies needed to detect one necessary CPM is calculated by: $(1 - p_t)/p_t$.
hospital-based studies with reasonable calibration. PredictCBC-2.0 may be used to tailor clinical decision-making toward CPM or alternative preventive strategies, especially when genetic information is available.

Abbreviations
AUC: Area under the ROC curve; BC: Breast cancer; BCAC: Breast Cancer Association Consortium; BMI: Body mass index; CBC: Contralateral breast cancer; CI: Confidence interval; CPM: Contralateral preventive mastectomy; DCA: Decision curve analysis; ER: Estrogen receptor; HEBON: The Hereditary Breast and Ovarian Cancer Research Group Netherlands; HER2: Human epidermal growth receptor 2; MICE: Multiple imputation by chained equations; O/E: Observed/expected; NCR: Netherlands Cancer Registry; PI: Prediction interval; PR: Progesterone receptor; PRS: Polygenic risk score; sHR: Subdistribution hazard ratio; TRIPOD: Transparent reporting of a multivariable prediction model for improvement.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13058-022-01567-3.

Additional file 1. Supplementary methods also including the following tables and figures Table S2. List of BCAC studies (including ABCS source) with the corresponding country and geographic area. Table S4: Clinical utility of the Breast Cancer risk models (PredictCBC and PredictCBC-2.0B). Figure S1. Visual assessment of calibration through calibration plots in the internal–external cross-validation at 5 years for the PredictCBC-2.0A model. Figure S2. Visual assessment of calibration through calibration plots in the internal–external cross-validation at 10 years for the PredictCBC-2.0A model. Figure S3. Visual assessment of calibration through calibration plots in the internal–external cross-validation at 10 years for the PredictCBC-2.0B model. Figure S4. Visual assessment of calibration through calibration plots in the internal–external cross-validation at 5 years for the PredictCBC-2.0B model. Figure S5. Density distribution of 5-year predicted contralateral breast cancer using PredictCBC-2.0 models. Figure S6. Decision curve analysis at 5 years for the contralateral breast cancer risk models (PredictCBC and PredictCBC-2.0) including BRCa mutation information.

Additional file 2: Table S1. Description of the studies included in the analyses.

Additional file 1: Table S3. Patient and primary breast cancer characteristics per study.

Acknowledgements
We thank all individuals who took part in these studies and all researchers, clinicians, technicians, and administrative staff who have enabled this work to be carried out. ABCFS thanks Maggie Angelakos, Judi Maskiell, and Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. ABCTB Investigators: Christine Clarke, Deborah Marsh, Rodney Scott, Robert Baxter, Desmond Yip, Jane Carpenter, Alison Davis, Nirmala Pathmanathan, Peter Simpson, J. Dinny Graham, Mythily Satchithanathan. ABCS and BOSOM thank all the collaborating hospitals and pathology departments and many individuals that made this study possible; specifically, we wish to acknowledge: Annegien Broeks, Sten Cornelissen, Frans Hogervorst, Laura van ‘t Veen, Emiel Rutgers. EMC thanks J.C. Blom-Leenheer, P. J. Bos, C.M.G. Crepin, and M. van Vliet for their contribution to this study. Special thanks are due to Stefano Bottelli for his R programming support. We highly appreciate the help of Tom Hueting to translate PredictCBC-2.0 into an online tool.

Author contributions
MBG and MHJ conceived the study in collaboration with EWS and MH. DG performed the statistical analysis. DG, MKS, MJK, EWS, and MH interpreted the results and drafted the manuscript. All remaining authors contributed to the critical revision and editing of the final version of the manuscript for publication. All authors read and approved the final manuscript.

Funding
This work is supported by the Alpe d’HuZes/Dutch Cancer Society (KWF Kankerbestrijding) project 6253. BCAC is funded by Cancer Research UK [C1287/16563, C1287/A10118], the European Union’s Horizon 2020 Research and Innovation Programme (Grant Numbers 634935 and 633784 for BRIDGES and B-CAST, respectively), and by the European Community’s Seventh Framework Programme under grant agreement number 223175 (Grant Number HEALTH-F2-2009-223175 (COGES)). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation, or writing of the report. Additional funding for BCAC is provided via the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health. The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia), and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. The ABCS study was supported by the Dutch Cancer Society (Grants NKI 2007-3839, 2009-4363). The work of the BCAC was partly funded by ELAN-Fond of the University Hospital of Erlangen. BOSOM was supported by the Dutch Cancer Society Grant Numbers DCS-NKI 2001-2423, DCS-NKI 2007-3839, and DCSNKi2009-4363; the Cancer Genomics Initiative; and notary office Spier & Hazenberg for the coding procedure. The BREEast Oncology Galician Network (BREGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI10/0125; the European Community’s 7th Framework Programme and FEDER PI10/0125/00018/COFINANCIADO FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrausalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomédica Galicia Sur. Xerencia
de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Conselleria de Industria Programa Sectorial de Investigación Aplicada, PIMEC-IV-D e I+D+I Sura del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Conselleria de Industria de la Xunta de Galicia, Spain; Grant EC11-192, Fomento de la Investigación Clinica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Interconecta. Ministerio de Economía y Competitividad, Xunta de Galicia, Spain. The EMC was supported by grants from Alpe d’HuZes/Dutch Cancer Society NKI2013-6253 and from Pink Ribbon 2012.WO39.C143. The HECBS was financially supported by the Helsinki University Hospital Research Fund, the Finnish Cancer Society, and the Sigrid Juselius Foundation. The HEBCN study is supported by the Dutch Cancer Society grants NKI1988-1854, NKI2004-3088, NKI2007-3756, NKI 12535, the Netherlands Organisation of Scientific Research grant NWO 91109024, the Pink Ribbon Grants 110005 and 2014-187. WO76, the BBMRI Grant NWO 184.021.007/CP46, and the Transcan Grant JTC 2012 Cancer 12-054. Financial support for KARBARC was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee foundation and Bert von Kantzows foundation. The KARMA study was supported by Märtis and Hans Rausings Initiative Against Breast Cancer. LMCB is supported by the Stichting tegen Kanker."The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01K10402]. MEc was supported by NIH grants CA63464, CA54281, CA09758, CA132839 and CA164973. The ORIGO study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The Netherlands Cancer Registry is hosted by the Netherlands Comprehensive Cancer Organisation (IKNL) and financed by the Dutch Ministry of Health, Welfare and Sports. The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The POSH study is funded by Cancer Research UK (grants C1275/ A11699, C1275/C22524, C1275/A19187, C1275/A15956 and Breast Cancer Campaign 2011PR62, 2013PR044). SKIDKSFZ is supported by the DKFZ. The SZBCS was supported by Grant PBZ-KBN_122/P05/2004 and the program of the Ministry of Science and Higher Education under the name “Regional Initiative of Excellence” in 2019–2022 project number 002/RD/2018/19 amount of financing 12 000 000 PLN. UBSCS was supported by funding from National Cancer Institute (NCI) grant R01 CA163533 (to N.J. Camp) and the Women’s Cancer Center at the Huntsman Cancer Institute (HCI). Data collection for UBSCS was supported by the Utah Population Database, Intermountain Health care and the Utah Cancer Registry which is funded by the NCI’s SEER Program (HHSN261201800016E), the US Centers for Disease Control and Prevention’s National Program of Cancer Registries (NUSDP006320), with additional support from the University of Utah and Huntsman Cancer Foundation.

Availability of data and materials
The datasets analyzed during the current study are not publicly available due to the protection of participant privacy and confidentiality, and ownership of the contributing institutions, but may be made available in an anonymized form via the corresponding author on reasonable request and after approval of the involved institutions.

Declarations
Ethics approval and consent to participate
All studies were approved by the appropriate ethics and scientific review boards. All procedures performed in studies involving human participants were in accordance with the ethical standards of international, national, and institutional research committees and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni Van Leeuwenhoek Hospital, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. 2Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands. 3Department of Biomedicine, EURAC Research Affiliated Institute of the University of Lubeck, Bolzano, Italy. 4Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands. 5Brandenburg Medical School, Institute of Biostatistics and Registry Research, Neuruppin, Germany. 6Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. 7Department of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. 8Department of Oncology, Örebro University Hospital, Örebro, Sweden. 9Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. 10Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. 11Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 12Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. 13Department of Internal Medicine and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA. 14Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. 15Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands. 16Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. 17Faculty of Medicine, University of Southampton, Southampton, UK. 18Division of Hematology and Oncology, Department of Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA. 19Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg (FAU), Erlangen, Germany. 20Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, UK. 21Cancer Research UK Edinburgh Centre, The University of Edinburgh, Edinburgh, UK. 22Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. 23Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. 24Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. 25Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany. 26Melbourne School of Population and Global Health, Centre for Epidemiology and Biostatistics, The University of Melbourne, Melbourne, VIC, Australia. 27Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. 28Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland. 29Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni Van Leeuwenhoek Hospital, Amsterdam, The Netherlands. 30Department of Moleculcular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. 31Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden. 32Department of Medical Genetics, University of Copenhagen, Copenhagen, Denmark. 33Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset, Stockholm, Sweden. 34Moores Cancer Center, University of California San Diego, La Jolla, CA, USA. 35Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, USA. 36Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. 37Department of Oncology, Unives Multidisciplinary Breast Center, Leuven Cancer Institute, University Hospitals Leuven, Louven, Belgium. 38Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. 39Department of Research and Development, Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, The Netherlands. 40Department of Health Technology and Services Research, Technical Medical Centre, University of Twente, Enschede, The Netherlands. 41Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia. 42Department of Clinical Pathology, The University of Melbourne, Melbourne, VIC, Australia. 43Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia. 44Department of Genetics, University Medical Center Groningen, University Groningen, Groningen, The Netherlands. 45Clinical Genetics, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. 46Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. 47Cancer Epidemiology Group, University Cancer Center Hamburg (UCCCH), University
References

1. Chen Y, Thompson W, Semenciw R, Mao Y. Epidemiology of contralateral breast cancer. Cancer Epidemiol Biomarkers Prev. 1999;8(10):855–61.
2. Gao X, Fisher SG, Emami B. Risk of second primary cancer in the contralateral breast in women treated for early-stage breast cancer: a population-based study. Int J Radiat Oncol Biol Phys. 2003;56(4):1038–45.
3. Curtis RE, Ron E, Hankey BF, Hoover RN. New malignancies following breast cancer. In: New malignancies among cancer survivors: SEER Cancer Registries. 1973–2000; 181–205.
4. Yu GP, Schantz SP, Neugut AI, Zhang ZF. Incidences and trends of second cancers in female breast cancer patients: a fixed inception cohort-based analysis (United States). Cancer Causes Control. 2006;17(4):411–20.
5. Soenjomatarim I, Louwman WJ, Lemmens VE, de Vries E, Klokman WJ, Coebergh JW. Risks of second primary breast and uterine cancer following female breast cancer in the south of The Netherlands, 1972–2001. Eur J Cancer Prev. 2005;14(1):231–7.
6. Schaapveld M, Visser Q, Louwman WJ, Willemsen PH, de Vries EG, van der Graaf WT, Otter R, Coebergh JW, van Leeuwen FE. The impact of adjuvant therapy on contralateral breast cancer risk and the prognostic significance of contralateral breast cancer: a population based study in the Netherlands. Breast Cancer Res Treat. 2008;108(1):189–97.
7. Turtle TM, Habermann EB, Grund EH, Morris TJ, Vnigm BA. Increasing use of contralateral prophylactic mastectomy for breast cancer patients: a trend toward more aggressive surgical treatment. J Clin Oncol. 2007;25(33):5203–9.
8. Narod SA. Bilateral breast cancers. Nat Rev Clin Oncol. 2014;11(3):157–66.
9. Metcalfe K, Gershenson M, Ghadirian P, Lynch HT, Snyder C, Tung N, Kim-Sing C, Eisen A, Foulkes WD, Rosen B, et al. Contralateral mastectomy and survival after breast cancer in carriers of BRCA1 and BRCA2 mutations: retrospective analysis. BMJ. 2014;348:g276.
10. Xiong Z, Yang L, Deng G, Huang X, Li X, Xie X, Wang J, Shuang Z, Wang X. Patterns of occurrence and outcomes of contralateral breast cancer: analysis of SEER data. J Clin Med. 2018;7(6):133.
11. Wong SW, Freedman RA, Saras A, Aydogan F, Barry WT, Golshan M. Growing use of contralateral prophylactic mastectomy despite no improvement in long-term survival for invasive breast cancer. Ann Surg. 2017;265(3):581–9.
12. Murphy JA, Milner TD, O’Donoghue JM. Contralateral risk-reducing mastectomy in sporadic breast cancer. Lancet Oncol. 2013;14(7):e262-269.
13. Basu NN, Hodson J, Chatterjee S, Gandhi A, Wisely J, Harvey J, Highton L, Murphy J, Barnes N, Johnson R, et al. The Angelina Jolie effect: contralateral risk-reducing mastectomy trends in patients at increased risk of breast cancer. Sci Rep. 2021;11(1):2847.
14. Domchek SM. Risk-reducing mastectomy in BRCA1 and BRCA2 mutation carriers: a complex discussion. JAMA. 2019;321(1):27.
15. Giardiello D, Steyerberg EW, Hauptmann M, Adank MA, Akdeniz D, Blomqvist C, Bojesen SE, Bolla MK, Brinkhuis M, Chang-Claude J, et al. Prediction and clinical utility of a contralateral breast cancer risk model. Breast Cancer Res. 2019;21(1):144.
16. Basu NN, Ross GL, Evans DG, Barr L. The Manchester guidelines for contralateral risk-reducing mastectomy. World J Surg Oncol. 2015;13:237.
17. Chowdhury M, Euhus D, Onega T, Waisfiz Q, Hollestelle A, Schutte M, et al. Prediction model for individualized risk prediction of contralateral breast cancer. Breast Cancer Res Treat. 2017;161(1):153–60.
18. Chowdhury M, Euhus D, Arun B, Umbricht C, Biswas S, Choudhary P. Validation of a personalized risk prediction model for contralateral breast cancer. Breast Cancer Res Treat. 2018;170(2):415–23.
19. Weischer M, Nordstegaard BG, Pharaoh P, Bolla MK, Nevalinnia H, van der Veer LJ, Garcia-Closas M, Hopper JL, Hall P, Andrusil IL, et al. CHEK2*1100delC heterozygosity in women with breast cancer associated with early death, breast cancer-specific death, and increased risk of a second breast cancer. J Clin Oncol. 2012;30(35):4308–16.
20. Akdeniz D, Schmidt MK, Seynaeve CM, McCool D, Giardiello D, van den Broek AJ, Hauptmann M, Steyerberg EW, Hoening MJ. Risk factors for metachronous contralateral breast cancer: a systematic review and meta-analysis. Breast. 2019;44:1–14.
21. Robson ME, Reiner AS, Brooks JD, Concannon PJ, John EM, Mellemkjaer L, Bernstein L, Malone KE, Knight JA, Lynch CF; et al. Association of common genetic variants with contralateral breast cancer risk in the WECARE study. J Natl Cancer Inst. 2017. https://doi.org/10.1093/jnci/djx051.
22. Fanale D, Incorvaia L, Filorizzo C, Bono M, Fierino A, Calo V, Brando C, Corsini LR, Barraco N, Badalament M, et al. Detection of germline mutations in a cohort of 139 patients with bilateral breast cancer by multi- gene panel testing: impact of pathogenic variants in other genes beyond BRCA1/2. Cancers (Basel). 2020;12(9):2415.
23. Kramer I, Hoening MJ, Mavaddat N, Hauptmann M, Keeman R, Steyerberg EW, Giardiello D, Antionio AC, Pharaoh PDP, Canisius S, et al. Breast cancer polygenic risk score and contralateral breast cancer risk. Am J Hum Genet. 2020;107(5):857–48.
24. Lakemen IMW, van den Broek AJ, Vos J, James DR, Asladi J, Andrusil IL, Arason A, Arnold N, Arun BK, Balmana J, et al. The predictive ability of the 313 variant-based polygenic risk score for contralateral breast cancer risk prediction in women of European ancestry with a heterozygous BRCA1 or BRCA2 pathogenic variant. Genet Med. 2021;23:1726–37.
25. Mavaddat N, Michioldoud K, Dennis J, Lush M, Fuchal L, Lee A, Tyrer JP, Chen TH, Wang Q, Bolla MK, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet. 2019;104(1):21–34.
26. Akdeniz D, Klaer MW, Smith CZA, Koppert LB, Hoening MJ. The impact of lifestyle and reproductive factors on the risk of a second primary cancer in the contralateral breast: a systematic review and meta-analysis. Cancer Causes Control. 2020;31(5):403–16.
27. Pijpe A, Monders B, Brohet RM, Collee JM, Verhoef S, Vasen HF, Hoogerbrugge X. Patterns of occurrence and outcomes of contralateral breast cancer: analysis of SEER data. J Clin Med. 2018;7(6):133.
28. Rieger PH, van Veen EB. Biobanking residual tissues. Hum Genet. 2011;130(3):357–68.
29. Foundation Federation of Dutch Medical Scientific Societies. Human tissue and medical research: code of conduct for responsible use. 2011.
30. van den Broek AJ, Schmidt MK, van’t Veer LJ, Oldenburg HSA, Rutgers EJ, Russell NS, Smit V, Voogd AC, Koppert LB, Siesling S, et al. Prognostic impact of breast-conserving therapy versus mastectomy of BRCA1/2 mutation carriers compared with noncarriers in a consecutive series of young breast cancer patients. Ann Surg. 2019;270(2):364–72.
31. Buuren S. Flexible imputation of missing data. Boca Raton: CRC Press; 2012.
32. Resche-Rigon M, White IR, Bartlett JW, Peters SA, Thompson SG. Group P-I5: Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. Stat Med. 2013;32(28):4890–905.
33. Van Buuren S. Flexible imputation of missing data. 2nd ed. Boca Raton: Chapman and Hall/CRC; 2018.
34. Geskus RB. Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. Biomet- rics. 2011;67(1):39–49.
35. Schoenfield DA. Sample-size formula for the proportional-hazards regression model. Biometrics. 1982;38(2):499–503.
36. Schmidt MK, Tollenaar RA, de Kemp SR, Broeks A, Cornelisse CJ, Smit VT, Petersen JL, van Leeuwen FE, van Veer LJ. Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation. J Clin Oncol. 2007;25(1):64–9.
37. Schmidt MK, Hogervorst F, van Hien R, Cornelissen S, Broeks A, Adank MA, Meijers H, Waisfiz Q, Hollestelle A, Schutte M, et al. Age- and tumor-subtype-specific breast cancer risk estimates for CHEK2*1100delC carriers. J Clin Oncol. 2016;34(23):2750–60.
38. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, external-external, and external validation. J Clin Epidemiol. 2016;69:245–7.
39. Austin PC, van Klaveren D, Vergouw Y, Nieboer D, Lee DS, Steyerberg EW. Geographic and temporal validity of prediction models: different approaches were useful to examine model performance. J Clin Epide- miol. 2016;79:76–85.
40. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med. 2016;35(2):214–26.

41. Blanche P, Dartigues JF, Jacquin-Gadda H. Estimating and comparing time–dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med. 2013;32(30):5381–97.

42. Brentnall AR, Cuzick J. Risk models for breast cancer and their validation. Stat Sci. 2020;35(1):1–30.

43. Austin PC, Putter H, Giardiello D, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICl) for competing risk models. Diagn Progn Res. 2022;6(1):2.

44. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of prediction models: individual patient data meta-analysis and network meta-analysis. Ann Intern Med. 2015;162(10):733–6.

45. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74.

46. Kerr KF, Brown MJ, 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(21):2534–40.

47. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008;8:53.

48. Heemskerk-Kersten BA, Roekus MA, Aarts GM, Aussem MG, Collie JM, Jansen L, Kets CM, Keymeulen KB, Koppert LB, Meijers-Heijboer HE, et al. Improved overall survival after contralateral risk-reducing mastectomy in BRCA1/2 mutation carriers with a history of unilateral breast cancer: a prospective analysis. Int J Cancer. 2015;136(6):688–77.

49. Balmiana J, Diez O, Rubio IF, Cardoso F, Group EGW. BRCA in breast cancer: ESMAO clinical practice guidelines. Ann Oncol. 2011;22(Suppl 6):S31–4.

50. Rutgers EJT. Is prophylactic mastectomy justified in women without BRCA mutation? Breast. 2019;48(Suppl 11):S52–6.

51. Giardiello D, Hauptmann M, Steyerberg EW, Adank MA, Akdeniz D, Blom JC, Blomqvist C, Bojesen SE, Bolla MK, Brinkhuis M, et al. Prediction of contralateral breast cancer: external validation of risk calculators in 20 international cohorts. Breast Cancer Res Treat. 2020;2018(12):423–34.

52. Antoniou AC, Pharoah PP, Smith R, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br J Cancer. 2004;91(8):1580–90.

53. Antoniou AC, Cunningham AP, Petro J, Evans DG, Laloo F, Nareo SA, Risch HA, Eyfjord JE, Hopper JL, Southey MC, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. Br J Cancer. 2008;98(9):1457–66.

54. Lee AJ, Cunningham AP, Tischkowitz M, Simard J, Pharoah PD, Easton DF, Antoniou AC. Incorporating truncating variants in PALB2, CHEK2, ATM into the BOADICEA breast cancer risk model. Genet Med. 2008;10(8):486–92.

55. Carver T, Hartley S, Lee A, Cunningham AP, Archer S, Babb de Villiers C, Roberts J, Ruston R, Walter FM, Tischkowitz M, et al. CanRisk Tool-A web interface for the prediction of breast and ovarian cancer risk and the likelihood of carrying genetic pathogenic variants. Cancer Epidemiol Biomarkers Prev. 2021;30(3):469–73.

56. Kramer I, Schaapveld M, Oldenburg HSA, Sonke GS, McCool D, van Leeuwen FE, Van de Vijver KK, Russell NS, Linn SC, Sesling S, et al. The influence of adjuvant systemic regimens on contralateral breast cancer risk and receptor subtype. J Natl Cancer Inst. 2019;111(7):709–18.

57. Witteveen A, Vliegen JM, Sonke GS, Klaase JM, Sesling S. Personalisation of breast cancer follow-up: a time-dependent prognostic nomogram for the estimation of annual risk of locoregional recurrence in early breast cancer patients. Breast Cancer Res Treat. 2015;152(5):627–36.

58. Volkel V, Hueting TA, Draeger T, van Maaren MC, de Munck L, Stroobbe LIA, Sonke GS, Schmidt MK, van Heezikwijk M, Groothuis-Oudshoorn CGM, et al. Improved risk estimation of locoregional recurrence, secondary contralateral tumors and distant metastases in early breast cancer: the INFLUENCE 2.0 model. Breast Cancer Res Treat. 2021;189:817–26.

59. Nieboer D, Vergouw Y, Ankerst DP, Roobol MJ, Steyerberg EW. Improving prediction models with new markers: a comparison of updating strategies. BMC Med Res Methodol. 2016;16(1):128.

60. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol. 2019;110:663–73.

61. Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ. 2012;344:e1681.

62. Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C, Wahlstrom C, Pooley KA, Parsons MT, Fortuno C, et al. Breast cancer risk genes—association analysis in more than 113,000 women. N Engl J Med. 2021;384(5):428–39.

63. Ho WK, Tan MM, Mavaddat N, Tai MC, Mariapun S, Li Y, Ho PJ, Dennis J, Tyer JP, Bolla MK, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. Nat Commun. 2020;11(3):3833.

64. Evans DG, van Veen EM, Byers H, Roberts E, Howell A, Howell SJ, Harkness EF, Brentnall A, Cuzick J, Newman WG. The importance of ethnicity: Are breast cancer polygenic risk scores ready for women who are not of White European origin? Int J Cancer. 2021;150:73–9.

65. Christodoulou E, Ma J, Collins GS, Stieberberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12–22.

66. Giardiello D, Antoniou AC, Mariani L, Easton DF, Steyerberg EW. Letter to the editor: a response to Ming’s study on machine learning techniques for personalized breast cancer risk prediction. Breast Cancer Res. 2020;22(1):17.

67. Thompson D, Easton D. The genetic epidemiology of breast cancer genes. J Mammary Gland Biol Neoplasia. 2004;9(3):221–36.

68. Reiner AS, Sisti J, John EM, Lynch CF, Brooks JD, Mellemkjaer L, Boice JD, Knight JA, Concannon P, Capanu M, et al. Breast cancer family history and contralateral breast cancer risk in young women: an update from the women’s environmental cancer and radiation epidemiology study. J Clin Oncol. 2018;36(15):1513–20.

69. Tokamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19(9):581–90.

70. Wald NJ, Old R. The illusion of polygenic disease risk prediction. Genet Med. 2019;21:1705–7.

71. Knight JA, Blackmore KM, Fan J, Malone KE, John EM, Lynch CF, Vachon CM, Bernstein L, Brooks JD, Reiner AS, et al. The association of mammographic density with risk of contralateral breast cancer and change in density with treatment in the WECARE study. Breast Cancer Res. 2018;20(1):23.

72. Van Belle V, Van Calster B. Visualizing risk prediction models. PLoS ONE. 2015;10(7):e0132614.

73. Bonnett LJ, Snell KE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. BMJ. 2019;365:l3737.

74. PREDICTCBC 2.0. https://www.evidencer.com/models/show/2949

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.