Double trouble: exploring risk factors to better predict contralateral breast cancer

Kramer, I.

CHAPTER 2

Breast cancer polygenic risk score and contralateral breast cancer risk

Am J Hum Genet. 2020 Nov 5;107(5):837-848

Iris Kramer
Maartje J. Hooning
Nasim Mavaddat
Michael Hauptmann
Renske Keeman
Ewout W. Steyerberg
Daniele Giardiello
Antonis C. Antoniou
Paul D.P. Pharoah
Sander Canisius
BCAC collaborators*
Douglas F. Easton
Per Hall
Marjanka K. Schmidt

**For full author list see publication*

Abstract

Previous research has shown that polygenic risk scores (PRS) can be used to stratify women according to their risk of developing primary invasive breast cancer. This study aimed to evaluate the association between a recently validated PRS of 313 germline variants (PRS_{313}) and contralateral breast cancer (CBC) risk. We included 56,068 women of European ancestry diagnosed with first invasive breast cancer from 1990 onwards with follow-up from the Breast Cancer Association Consortium. Metachronous CBC risk (N=1,027) according to the distribution of the PRS_{313} was quantified using Cox regression analyses. We assessed PRS_{313} interaction with age at first diagnosis, family history, morphology, ER-, PR-, and HER2-status, and (neo)adjuvant therapy. In Asian studies, with limited follow-up, CBC risk associated with PRS_{313} was assessed using logistic regression for 340 women with CBC compared with 12,133 women with unilateral breast cancer. Higher PRS_{313} was associated with increased CBC risk: hazard ratio per standard deviation (SD)=1.25 (95%CI=1.18-1.33) for Europeans, and an OR per SD=1.15 (95%CI=1.02-1.29) for Asians. The absolute lifetime risks of CBC, accounting for death as competing risk, were 12.4% for European women at the 10th percentile and 20.5% at the 90th percentile of the PRS_{313}. We found no evidence of confounding by, or interaction with patient characteristics, characteristics of the primary tumor, or treatment. The C-index for the PRS_{313} alone was 0.563 (95%CI=0.547-0.586). In conclusion, the PRS_{313} is an independent factor associated with CBC risk, and may be incorporated in CBC risk prediction models to help improve stratification of patients and optimize surveillance and treatment strategies.

Introduction

Due to the high incidence of breast cancer and improving survival, an increasing number of breast cancer survivors are at risk of developing contralateral breast cancer (CBC). The 10-year cumulative incidence of CBC is ~4%[1,2], however estimates vary widely depending on factors such as germline genetics, family history, and (neo)adjuvant systemic therapy for the first breast cancer[3]. The risk of developing CBC is particularly high in women with rare mutations in certain genes including *BRCA1, BRCA2,* and *CHEK2*, with approximately two- to fourfold higher risks reported compared with women without these mutations[3].

Recently, genome-wide association studies (GWAS) have identified multiple common germline variants that are associated with first primary breast cancer risk[4,5]. These are associated with small differences in risk individually, but their combined effects can be summarized in a polygenic risk score (PRS), which has been shown to stratify women according to their risk of developing breast cancer[6-9]. Using a large GWAS dataset from the Breast Cancer Association Consortium (BCAC), we previously developed and validated a 313-variant PRS (PRS_{313}) among women of European descent. In independent prospective studies, this PRS_{313} predicted the risk of primary invasive breast cancer with an odds ratio (OR) per standard deviation (SD) of 1.61 (95% confidence interval (95%CI)=1.57-1.65)[7]. The PRS_{313} has also been externally validated using the UK Biobank cohort.

The aim of the current study was to evaluate the association between PRS_{313} and CBC risk, using data from BCAC. Other studies have shown associations between risk of CBC and both a 67-variant PRS[10] and individual variants[11], but not yet with PRS_{313}, the most extensively validated PRS. Further, the dataset currently evaluated is larger than those previously tested. We carried out two types of analyses. We conducted a cohort study among studies of European ancestry women with follow-up data available, and performed Cox regression analyses to estimate hazard ratios (HRs) for CBC. Potential confounding and interaction with patient characteristics, characteristics of the primary tumor, or treatment were tested. In addition, to directly compare with the OR reported for PRS_{313} and first breast cancer, we selected case-case series and performed logistic regression analyses comparing the PRS_{313} distribution in women with CBC versus those with unilateral breast cancer. These analyses were conducted separately in European and Asian women (follow-up was too limited to perform a cohort study for the Asian population). Use of PRS_{313} may lead to more accurate CBC risk prediction to support decision making for women who may or may not benefit from additional surveillance and risk-reducing treatment strategies.

Material and Methods

Study subjects

Case-case series

We selected women who were diagnosed with breast cancer and women without any diagnosis of breast cancer from the BCAC including all women of European ancestry, based on genotyping data, selecting only those studies which reported on CBC (62 studies) (Figure S1A, Table S1-S2). BCAC database version freeze 12 was used. All women diagnosed with invasive breast cancer as a first cancer were included in the analysis; the small number of tumors with unknown invasiveness were considered invasive (Table S2). In the case-case series, a CBC was defined as a breast cancer (in situ or invasive) in the contralateral breast irrespective of the time since the first breast cancer. The case-case series comprised 81,000 women with unilateral breast cancer, 3,607 women with CBC, and 62,830 women without any diagnosis of breast cancer (Figure S1A). We also compared women with unilateral breast cancer to women without any diagnosis of breast cancer to reproduce the estimate that was previously reported for first breast cancer risk[7] in our study selection.

We selected for a separate analysis women of Asian ancestry of the BCAC data comprising 12,133 women with unilateral breast cancer, 340 women with CBC, and 13,398 women without any diagnosis of breast cancer from eight studies (Figure S1B, Table S2).

European cohort

In the European cohort we used metachronous CBC as the outcome, defined as a breast cancer in the contralateral breast (in situ or invasive) diagnosed at least three months after the first breast cancer. We used a cut-off of three months to reduce the likelihood that these CBCs represent metastases rather than true second primary tumors. We selected all women diagnosed with breast cancer from the European case-case series and excluded four studies that did not provide follow-up information on vital status (Figure S1A). We did not include Asian women since follow-up was too limited in these studies. We additionally excluded 6,207 women with no follow-up and 2,208 women who developed synchronous CBC, distant metastasis, or who died or last known to be alive within three months after the first breast cancer diagnosis. Since BCAC also included prevalent cases, we excluded 3,796 women who developed CBC or were censored before study entry. The case-case series included women diagnosed between 1947 and 2018. In the European cohort, we excluded 2,235 women who were diagnosed with their first breast cancer before 1990 or who had missing year of first diagnosis. We restricted to women diagnosed from 1990 onwards so that diagnostic procedures and treatment would be more representative of current practice. Moreover, clinico-pathological, treatment and follow-up data were more complete after 1990. In addition, we excluded 16 studies (9,783women) without information about metachronous CBC events (Figure S1A). After these exclusions, the cohort for this analysis comprised data from 42 studies, including 56,068 women with invasive breast cancer among whom 1,027 metachronous CBC occurred (Table S2).

All individuals provided written informed consent, and all studies were approved by the relevant institutional review boards. BCAC data were centrally harmonized and cleaned in communication with the study data managers and principal investigators. Data collection for individual studies is described in Table S1.

Genotyping and PRS

DNA samples from participants were genotyped using the iCOGS array[12,13] or the OncoArray[4,14], with genotypes for variants not on the arrays estimated by imputation[4,13]. The PRS_{313} was calculated as a weighted sum of the minor allele dosages; the variant selection and weights are as given by Mavaddat et al.[7]. We also calculated estimates for a previously published PRS_{77}[6], and estrogen receptor (ER)-specific PRSs (ER-positive PRS_{313} and ER-negative PRS_{313})[7]. The ER-specific PRSs were constructed by defining subtype-specific weights for the 313 variants using a hybrid approach[7]. Variants and corresponding coefficients used to construct the PRS are shown in Table S3. We standardized the PRS in our analyses by dividing it by the SD of the PRS of the controls (PRS_{77} SD=0.45; PRS_{313} SD=0.61; ER-positive PRS_{313} SD=0.65; ER-negative PRS_{313} SD=0.59) exactly as was done in the analyses of the PRS and first breast cancer risk[6,7]. This allows a direct comparison of the magnitude of the CBC relative risk estimation to that of the first breast cancer.

For samples genotyped with both OncoArray and iCOGS array (9,071 samples), OncoArray data were used in preference as the imputation quality was generally higher. The intraclass correlation coefficient (ICC) between the PRS derived from the two platforms was 0.99 (95%CI=0.99-0.99) for the PRS_{77}, and 0.96 (95%CI=0.95-0.96) for PRS_{313} (Figure S2). Given the high correlation between the two platforms, PRS measures from both platforms were used in the analyses without adjustment.

Statistical analysis

European cohort

The primary outcome in the European cohort was the development of metachronous CBC. Cox proportional hazards models were used to estimate HRs for metachronous CBC risk by PRS, stratified by country. Since previous studies have shown that age at first breast cancer diagnosis is an important predictor of CBC[3], the analyses were performed with attained age as the time scale. Time at risk started three months after the first breast cancer diagnosis and ended at the age of CBC diagnosis, distant metastasis (where available), death, or end of follow-up, whichever came first. For patients that had a study entry more than three months after first breast cancer diagnosis, follow-

up started at the age of study entry. We also performed a fixed-effect meta-analysis of country-specific effects using the STATA command *metan*. We performed a fixed-effect meta-analysis over a random-effect meta-analysis since there was no evidence for heterogeneity in effect sizes between countries (I-squared=0%, Figure S3). For some analyses, only invasive CBC was used as the outcome; in these analyses we censored on in situ CBC. Separate analyses were conducted for ER-positive CBC (censored on ER-negative- and ER-unknown CBC) and ER-negative CBC (censored on ER-positive- and ER-unknown CBC).

We evaluated the linearity of the association between PRS_{313} per unit SD and CBC risk using restricted cubic splines with three knots. There was no evidence for violation of the linearity assumption. Therefore, in the main analysis, the PRS_{313} was treated as a continuous covariate, and estimated the HR per unit SD of the $PRS_{313.}$ Violation of the proportional hazard assumption was assessed by inspection of the Schoenfeld residuals[15]. As a second analysis, we used the per SD log HR of the PRS_{313} to calculate the predicted HR at different percentiles of the PRS_{313}, compared to the 50th percentile. Third, the PRS_{313} was categorized into percentile groups (0th to 10th, 10th to 20th, 20th to 40th, 40th to 60th, 60th to 80th, 80th to 90th, 90th to 100th) to illustrate the differences between PRS_{313} subgroups, with the middle quintile (40th to 60th) as the reference.

We also performed multivariable Cox regression analyses to determine whether the log HR of CBC risk by PRS changed when adjusting for year of first breast cancer diagnosis, family history of breast cancer in a first degree relative, and several clinical characteristics of the first breast cancer such as nodal status, tumor size, morphology, ER-, progesterone receptor (PR)- and human epidermal growth factor receptor 2 (HER2)-status, (neo)adjuvant chemotherapy, adjuvant endocrine therapy, and radiotherapy. These analyses were performed in all patients, a complete case set (excluding patients with unknown values for the covariates), and in a set excluding studies oversampling cases with family history. Potential effect modification of the PRS_{313} effect by the same variables was evaluated by fitting interaction terms in different models using complete case sets, including the standardized PRS_{313}, modifier, and interaction.

The discriminative ability of different models; ([model 1] PRS_{313} alone, [model 2] other risk factors (the adjustment variables from the multivariable Cox regression analyses), [model 3] PRS_{313} + other risk factors) was calculated using Harrell's C-index[16]. Since no standard performance measures are currently available to account for left-truncated follow-up time (*i.e.*, to start analyses at age at study entry), we used time since first breast cancer as the time scale to calculate the C-index.

Absolute risks

Absolute risks of developing CBC at PRS_{313} percentiles were calculated using the estimated log HRs per SD from the breast cancer cohort (BCAC) under the log-linear model, assuming the PRS is normally distributed. The PRS_{313}- and age-specific incidences were constrained to the age-specific CBC incidences from women diagnosed with a first invasive breast cancer in the period 2003-2010 from the Netherlands Cancer Registry (NCR)[1]. The procedure for constraining the incidences has been previously described[17]. The age-specific CBC incidences were calculated overall and for age-specific groups, censoring on death and distant metastasis. We used data from the NCR since this registry has complete coverage of all newly diagnosed cancers in the Netherlands. The NCR cohort included all females aged ≥18 years and follow-up for second cancers was complete until February 1, 2016[1]. We then applied the competing risk of dying on the absolute CBC risks. The absolute CBC risk (AR_g) by age t in PRS_{313} category g, taking into account the competing risk of dying was calculated by:

$$AR_g(t) = \sum_{u=0}^{t-1} \mu_g(u) S_g(u) S_m(u)$$

Where $\mu_g\ (t)$ is the CBC incidence associated with PRS_{313} category g, $S_g\ (t)$ the probability of being free of CBC to age t, and $S_m\ (t)$ the probability of surviving to age t.

Case-case series

For the case-case series (European and Asian), logistic regression models were used to estimate the ORs for CBC risk (comparing with unilateral breast cancer) and for unilateral breast cancer risk (comparing with women without any diagnosis of breast cancer) associated with PRS_{313}. All analyses were adjusted for age and country (Table S1). For all unilateral- and contralateral breast cancer patients we used age at first breast cancer diagnosis, and for women without any diagnosis of breast cancer we used age at baseline questionnaire.

For direct comparison with the estimate reported for PRS_{313} and first breast cancer, we also performed logistic regression analyses in the same BCAC study participants included in the validation of the association between PRS_{313} and first breast cancer risk[7]. This validation set comprised a subsample from 24 studies and included 3,781 women with unilateral breast cancer, 94 women with CBC, and 3,753 women without any diagnosis of breast cancer (Table S2). For this analysis, we adjusted for 10 principal components, in line with Mavaddat et al.[7].

For European women who had follow-up time available more than three months after the first breast cancer diagnosis, a sensitivity analysis was performed for metachronous CBC (1,702 CBCs). We also did a separate analysis for invasive CBC (N=3,246), by excluding CBC in situ.

All P-values are two sided; tests with P<.05 are referred to as statistically significant. Analyses were performed using STATA, version 13.1 (StataCorp) and R version 3.3.2.

Results

European (cohort) Cox regression analyses

The European cohort included 56,068 women diagnosed with first invasive breast cancer with 1,027 metachronous CBC events. Median follow-up was 8.4 years. Patient, tumor, and treatment characteristics are summarized in Table S4.

The associations between the different PRSs and CBC risk are shown in Table 1. The HR for CBC per SD of PRS_{313} was 1.25 (95%CI=1.18-1.33). For comparison, the HR per SD for PRS_{77} was 1.21 (95%CI=1.14-1.29). Women within the 0th to 10th and the 90th to 100th percentile of the PRS_{313} had 0.59-fold (95%CI=0.45-0.78) and 1.38-fold (95%CI=1.13-1.69) risks of CBC, respectively, compared with women within the 40th to 60th percentile (Figure 1, Table S5). The predicted HRs of CBC for women at the 10th and 90th percentile of the PRS_{313} were 0.75 and 1.33, respectively, compared to the 50th percentile (Figure 1). Since we observed evidence of departure from the proportional hazards assumption (P=0.02)[15], we also calculated HRs stratified for follow-up duration (<five and ≥five years). The HR by SD of the PRS_{313} was 1.21 (95%CI=1.10-1.32) for CBC diagnosed ≤five years after first breast cancer diagnosis (CBC N=428), and 1.28 (95%CI=1.18-1.38) for CBC diagnosed >five years after first diagnosis (CBC N=599).

Table 1. Association between PRSs and contralateral breast cancer risk in the European cohort (N=56,068)

Polygenic risk score (PRS)		No. of CBC	HR per unit SD[a]	95%CI	P-value
PRS_{77} [b]					
	All CBC	1,027	1.21	1.14-1.29	<.001
	Invasive CBC	923	1.21	1.13-1.29	<.001
PRS_{313} [b]					
	All CBC	1,027	1.25	1.18-1.33	<.001
	Invasive CBC	923	1.24	1.16-1.32	<.001
	ER-positive invasive CBC[d]	275	1.38	1.23-1.55	<.001
	ER-negative invasive CBC[d]	97	0.92	0.75-1.12	.39
ER-positive PRS_{313} [b,c]					
	All CBC	1,027	1.23	1.16-1.31	<.001
	Invasive CBC	923	1.22	1.15-1.30	<.001
	ER-positive invasive CBC[d]	275	1.37	1.22-1.54	<.001
ER-negative PRS_{313} [b,c]					
	All CBC	1,027	1.25	1.17-1.33	<.001
	Invasive CBC	923	1.24	1.16-1.33	<.001
	ER-negative invasive CBC[d]	97	1.06	0.86-1.30	.58

Abbreviations: PRS = polygenic risk score; No. = number; CBC = contralateral breast cancer; HR = hazard ratio; CI = confidence interval; ER = estrogen receptor; SD = standard deviation

[a] All analyses were performed with attained age as time scale

[b] Coefficients to construct the PRSs are shown in Table S3. All PRSs were standardized by the same SD as was used by Mavaddat et al.[7]. The SD was 0.45 for overall breast cancer PRS_{77}, 0.61 for overall breast cancer PRS_{313}, 0.65 for ER-positive PRS_{313}, and 0.59 for ER-negative PRS_{313}

[c] ER-specific PRSs were constructed using a hybrid method, as described by Mavaddat et al.[7]

[d] Patients with ER-unknown CBC (N=551) were censored in these analyses

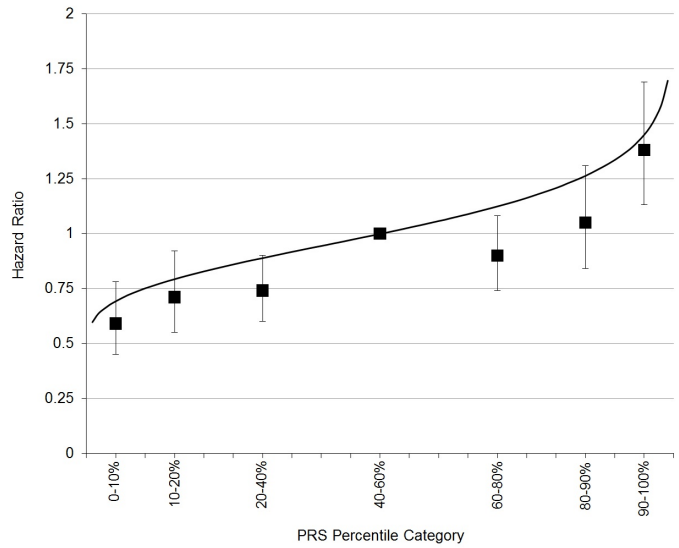

Figure 1. Estimates for contralateral breast cancer risk by percentile categories of the 313-variant PRS (PRS_{313})

The figure shows the hazard ratios per SD and 95% confidence intervals for percentiles of the PRS_{313} relative to the middle quintile (underlying table can be found in Table S5). The solid line denotes the estimates for contralateral breast cancer risk with the PRS_{313} fitted as a continuous covariate. Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al.[7]. The analyses were performed with attained age as time scale. Abbreviations: PRS = polygenic risk score; SD = standard deviation

The HR per SD of PRS_{313} for ER-positive invasive CBC was 1.38 (95%CI=1.23-1.55), compared to a HR per SD of the ER-positive PRS_{313} of 1.37 (95%CI=1.22-1.54) (Table 1). For ER-negative invasive CBC, the HR per SD was 0.92 (95%CI=0.75-1.12) for PRS_{313} and 1.06 (95%CI=0.86-1.30) for the ER-negative PRS_{313}.

Sensitivity analysis using the overall PRS_{313} showed a HR per SD of 1.24 (95%CI=1.16-1.32) for invasive CBC risk. When we used time since first breast cancer as the time scale, we found similar results (HR per SD=1.25, 95%CI=1.18-1.33). Meta-analysis of country-specific effects showed a HR per SD of 1.25 (95%CI=1.18-1.33) for CBC risk by PRS_{313} (Figure S3).

The association between the PRS_{313} and CBC risk did not change when adjusting for patient, tumor, and treatment characteristics, nor when excluding studies oversampling cases with a family history (Table S6). When considering potential modifiers of the effect of the PRS_{313} on CBC risk (Table 2), we found that the HR was the lowest in women aged <40 years at first breast cancer diagnosis (HR per SD=1.13; 95%CI=0.98-1.31), and tended to increase with age, although these effects were not statistically significant ($P_{heterogeneity}$=.26; P_{trend}=.05). We found no indication for effect modification by family history ($P_{heterogeneity}$=.63), morphology ($P_{heterogeneity}$=.14), ER-status ($P_{heterogeneity}$=.13), PR-status (P=.26), HER2-status ($P_{heterogeneity}$=.42), chemotherapy ($P_{heterogeneity}$=.60), endocrine therapy ($P_{heterogeneity}$=.79), or radiotherapy ($P_{heterogeneity}$ =.40) (Table 2).

Table 2. Association between the 313-variant PRS (PRS_{313}) and contralateral breast cancer risk for subgroups

Subgroups	No. of patients	No. of CBC	HR per unit SD[a,b]	95%CI	P-value	$P_{hetero-geneity}$[c,d]	P_{trend}[c,e]
All patients	56,068	1,027	1.25	1.18-1.33	<.001	-	-
Age at first breast cancer diagnosis (years)						.26	.05
<40	5,877	171	1.13	0.98-1.31	.09		
40-49	11,928	265	1.25	1.11-1.41	<.001		
50-59	16,882	320	1.22	1.09-1.36	<.001		
60+	21,381	271	1.36	1.21-1.52	<.001		
Family history (first degree relative)						.63	-
no	33,623	618	1.26	1.16-1.36	<.001		
yes	10,369	302	1.22	1.09-1.36	<.001		
Morphology						.14	-
ductal	37,324	621	1.21	1.12-1.31	<.001		
lobular	5,878	118	1.32	1.10-1.59	.002		
mixed (ductal and lobular)	2,174	46	1.52	1.15-2.02	.004		
other	3,344	70	1.20	0.96-1.50	.11		
ER-status						.13	-
negative	9,527	194	1.13	0.98-1.30	.08		
positive	38,090	670	1.28	1.19-1.38	<.001		
PR-status						.26	-
negative	13,098	244	1.16	1.03-1.32	.02		
positive	27,044	554	1.27	1.17-1.38	<.001		
HER2-status						.42	-
negative	23,787	352	1.29	1.17-1.44	<.001		
positive	4,969	60	1.45	1.13-1.85	.004		
(Neo)adjuvant chemotherapy						.60	-
no	18,110	361	1.28	1.16-1.42	<.001		
yes	18,559	363	1.24	1.12-1.37	<.001		
(Neo)adjuvant endocrine therapy						.79	-
no	10,781	242	1.28	1.13-1.44	<.001		
yes	27,322	460	1.30	1.19-1.43	<.001		
Radiotherapy						.40	-
no	11,023	188	1.33	1.15-1.53	<.001		
yes	29,142	617	1.24	1.15-1.34	<.001		

Abbreviations: PRS = polygenic risk score; No. = number; CBC = contralateral breast cancer; HR = hazard ratio; CI = confidence interval; ER = estrogen receptor; PR = progesterone receptor; HER2 = human epidermal growth factor receptor 2

[a] HR for CBC risk by unit SD of PRS_{313}. All analyses were performed with attained age as time scale

[b] Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by standard deviation=0.61, in line with Mavaddat et al.[7]

[c] The interaction between the PRS_{313} and each subgroup was tested in different models including the standardized PRS_{313}, modifier, and interaction. Patients with unknown values were excluded from these analyses. Since attained age was used as time scale in all models, the model with age at first breast cancer only included the PRS_{313} and interaction

[d] P for interaction based on test for heterogeneity across categories

[e] P for interaction based on a trend test with age as continuous variable

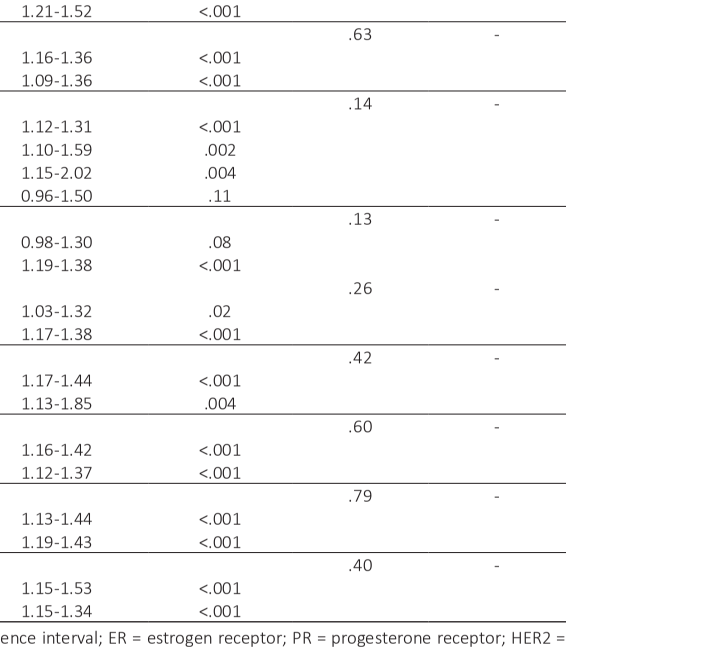

Figure 2. Predicted contralateral breast cancer risk by percentile of the 313-variant PRS (PRS_{313}) with death as competing risk

Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al.[7] The CBC incidences were calculated based on incidence data from the Netherlands Cancer Registry[8] and relative risks estimated as described in the Material and Methods. Abbreviations: PRS = polygenic risk score; CBC = contralateral breast cancer

The C-index was 0.563 (95%CI=0.547-0.586) for the model only including PRS_{313}, 0.605 (95%CI=0.591-0.629) for the model only including other risk factors, and 0.623 (95%CI=0.608-0.645) for the complete model (Table 3).

Absolute risks

Based on the HR estimates for PRS_{313}, the predicted CBC risk by age 80 years was 12.4% at the 10th percentile of the PRS_{313}, compared with 20.5% at the 90th percentile of the PRS_{313} (Figure 2), accounting for death as competing risk. When death was not taken into account as competing risk, the corresponding predicted risks by age 80 were 17.0% at the 10% percentile and 27.9% at the 90th percentile of the PRS_{313} (Figure S4). Table 4 shows the five- and 10-year cumulative CBC risks by PRS_{313} for different age groups, accounting for death as competing risk (Table S7 shows results without competing risks).

European and Asian (case-case series) logistic regression analyses

Figure 3 shows the distribution of the PRS_{313} per SD in the European case-case series. Median PRS_{313} was -0.4 (interquartile range [IQR]=1.35) for control women without any diagnosis of breast cancer (N=81,000), 0.2 (IQR=1.36) for women with unilateral breast cancer (N=62,830), and 0.5 (IQR=1.40) for women with CBC (N=3,607). The OR

for unilateral breast cancer per SD of the PRS_{313}, compared to control women, was 1.82 (95%CI=1.80-1.84) (Table S8). The OR for CBC per SD of PRS_{313}, compared to unilateral breast cancer, was 1.30 (95%CI=1.26-1.35) .

In sensitivity analyses, the OR per SD of PRS_{313} was 1.27 (95%CI=1.21-1.33) for metachronous CBC and the OR per SD was 1.29 (95%CI=1.24-1.33) for invasive CBC, compared to unilateral breast cancer. When analyses were restricted to the validation set of Mavaddat et al[7], the OR for unilateral breast cancer per SD of the PRS_{313} was 1.67 (95%CI=1.59-1.76) compared to control women, and the OR for CBC per SD of PRS_{313} was 1.39 (95%CI=1.13-1.70) compared to unilateral breast cancer (Table S8).

For women of Asian descent, the OR for unilateral breast cancer per SD of the PRS_{313} was 1.56 (95%CI=1.52-1.60) compared to control women, and the OR for CBC per SD of PRS_{313} was 1.15 (95%CI=1.02-1.29) compared to women with unilateral breast cancer (Table S8).

Table 3. Discriminatory ability (C-index) of the 313-variant PRS (PRS_{313}) and other risk factors for contralateral breast cancer risk in the European cohort

	C-index (95%CI)[a,b]
Model 1	
PRS_{313}[c] alone	0.563 (0.547-0.586)
Model 2	
Other risk factors[d]	0.605 (0.591-0.629)
Model 3	
PRS_{313}[c] + other risk factors[d]	0.623 (0.608-0.645)

Abbreviations: PRS = polygenic risk score; CI = confidence interval

[a] The Harrell's C-index was obtained by the STATA stcox postestimation command 'estat concordance', using time since first breast cancer on the time scale without taking delayed entry (prevalent cases) into account. We did not consider delayed-entry since no standard performance measures are currently available in the statistical literature to account for left-truncated follow-up time. The median of delayed entry was 0.4 years (standard deviation=2.7) in our study

[b] The 95% CIs were obtained by use of the 'somersd' package in STATA

[c] Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al.[7]

[d] Including age at first diagnosis, year of first diagnosis, family history for breast cancer in a first degree relative, and clinical characteristics of the first breast cancer (nodal status, tumor size, differentiation grade, morphology, estrogen receptor status, human epidermal growth factor receptor 2 status, chemotherapy, endocrine therapy, radiotherapy)

Table 4. Five- and ten-year cumulative risks of contralateral breast cancer by the 313-variant PRS (PRS_{313}) for different age groups with death as competing risk

	5-year cumulative CBC risks (%) range by age						10-year cumulative CBC risks (%) range by age					
Age at first breast cancer diagnosis (years)	5th percentile PRS_{313}	10th percentile PRS_{313}	50th percentile PRS_{313}	90th percentile PRS_{313}	95th percentile PRS_{313}		5th percentile PRS_{313}	10th percentile PRS_{313}	50th percentile PRS_{313}	90th percentile PRS_{313}	95th percentile PRS_{313}	
30-34	1.9-3.1	2.1-3.4	2.7-4.5	3.6-5.9	4.0-6.5		3.1-4.1	3.4-4.5	4.5-5.9	5.9-7.7	6.5-8.5	
35-39	0.8-2.1	0.9-2.3	1.2-3.0	1.5-3.9	1.7-4.3		2.1-3.5	2.3-3.8	3.0-5.0	3.9-6.6	4.3-7.2	
40-44	1.5-2.8	1.7-3.1	2.2-4.1	2.9-5.3	3.2-5.9		2.8-4.6	3.1-5.0	4.1-6.6	5.3-8.6	5.9-9.4	
45-49	1.4-2.5	1.5-2.7	2.0-3.6	2.6-4.7	2.9-5.2		2.5-3.9	2.7-4.3	3.6-5.6	4.7-7.4	5.2-8.1	
50-54	1.4-2.8	1.5-3.0	1.9-4.0	2.6-5.2	2.8-5.8		2.8-4.5	3.0-4.9	4.0-6.4	5.2-8.4	5.8-9.3	
55-59	1.6-3.1	1.8-3.4	2.3-4.5	3.1-5.9	3.4-6.5		3.1-4.8	3.4-5.2	4.5-6.9	5.9-9.0	6.5-9.9	
60-64	1.7-3.3	1.9-3.6	2.5-4.7	3.3-6.2	3.6-6.8		3.3-5.0	3.6-5.4	4.7-7.1	6.2-9.3	6.8-10.2	
65-70	1.5-3.2	1.6-3.5	2.1-4.6	2.8-6.1	3.1-6.7		3.2-4.1	3.5-4.5	4.6-5.9	6.1-7.7	6.7-8.5	

Abbreviations: PRS = polygenic risk score; CBC = contralateral breast cancer

Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al[7]. The CBC incidences for each age group were calculated based on incidence data from the Netherlands Cancer Registry[1] and relative risks estimated as described in the Material and Methods. Death was taken into account as competing risk

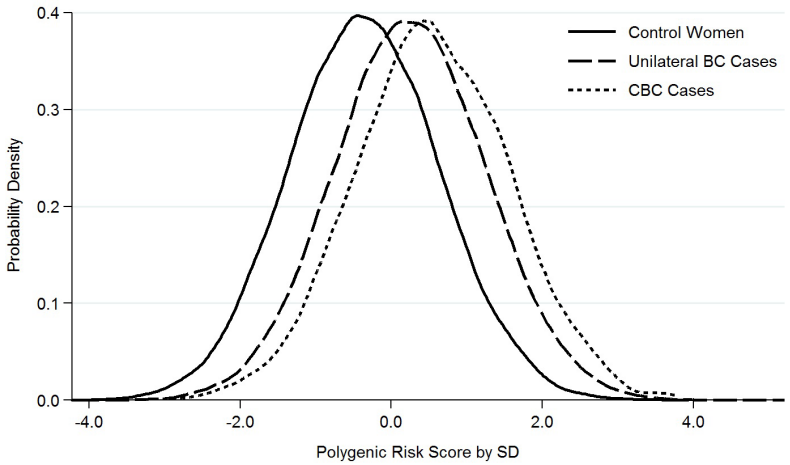

Figure 3. Distribution of the 313-variant PRS (PRS_{313}) in 62,830 control women without any diagnosis of breast cancer, 81,000 women with unilateral breast cancer, and 3,607 women with contralateral breast cancer

Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al.[7]. Abbreviations: BC = breast cancer; CBC = contralateral breast cancer; PRS = polygenic risk score; SD = standard deviation

Discussion

Previous studies have shown that a PRS, summarizing the effects of common germline variants, can be used to stratify women with respect to their risk to develop a primary breast cancer[6-9]. In this study, we observed a clear association between the PRS_{313} and CBC risk in women of both European and Asian ancestry. The association was observed in both the case-case series and the European cohort. The HRs per SD of CBC for women at the 10th and 90th percentile of the continuous predicted PRS_{313} were 0.75 and 1.33, respectively, compared to the 50th percentile. This translates to absolute risks at the 10th and the 90th percentile of the PRS_{313} of 12.4% and 20.5%, respectively, by age 80 years. We estimated a C-index for the PRS_{313}, summarizing its discriminatory ability, of 0.563 in the European cohort.

One previous study has investigated the effect of a PRS, including 67 variants, and CBC risk[10]. This study found a risk ratio of 1.75 (95%CI=1.41-2.18) for women in the upper quartile of the PRS compared with women in the lowest quartile. To facilitate comparison, we performed a similar analysis in our case-case series, showing an OR of 1.98 (95%CI=1.79-2.18), adjusted for country and age at first diagnosis, for women in the upper quartile of the PRS_{313}. This indicates the PRS_{313} improves stratification relative to PRSs including fewer variants. Moreover, in our European cohort, the C-index for the PRS alone improved from 0.547 (95%CI=0.536-0.575) for the previously reported PRS_{77}[6] to 0.563 (95%CI=0.547-0.586) for the PRS_{313}.

We found no evidence that the association between the PRS_{313} and CBC risk was confounded by family history, adjuvant therapy, morphology, age, or tumor receptor status of the first breast cancer, nor that there was effect modification by those factors. The absence of notable effect modification is in line with the abovementioned study of a 67-variant PRS and CBC risk; no heterogeneity in association was found by age, family history, morphology, ER-status, and adjuvant treatment[10].

To provide an external validation of our findings, we examined data from UK Biobank, which includes many women diagnosed with breast cancer with data available on the PRS_{313} (Supplemental Note). Unfortunately, UK Biobank has no information available on the laterality of the tumor, and it is, therefore, not possible to distinguish between contralateral and ipsilateral breast cancers. We therefore performed analyses using any second breast cancer as the endpoint. This secondary analysis did confirm the association between the PRS_{313} and second breast cancer risk (HR per SD=1.13, 95%CI=1.01-1.27), but with a lower estimate than in our European cohort. The lower estimate may be explained by the inclusion of the ipsilateral breast cancers, which may be more likely to be recurrences than new primary breast cancers compared to CBCs. Indeed, when we used ipsilateral breast cancer as the outcome in our European cohort, we found no association with the PRS_{313} (HR=1.02, 95%CI=0.90-1.15).

The association between the PRS_{313} and CBC risk (OR per SD=1.30; 95%CI=1.26-1.35) in the BCAC database was weaker (expressed in terms of an OR) than was found for first breast cancer among independent prospective studies (OR per SD=1.61; 95%CI=1.57-1.65). Under a simple polygenic model, the relative risk would be expected to be similar for the second breast cancer. The attenuated estimate for CBC might however be explained by several factors. Some attenuation of the estimate might have been due to dilution in the end-point definition, *i.e.*, if some of the CBCs were metastases. Previous studies investigating the clonal relatedness of first breast cancers and CBCs using tumor sequencing have shown that 6-12% of CBCs represent metastases[18,19]. This hypothesis would be consistent with our finding of a slightly stronger association between the PRS_{313} and late CBCs, diagnosed >five years after the first breast cancer, than for early CBCs, diagnosed ≤five years after the first cancer, since the latter are more likely to be metastases. In addition, 3-5% of the breast cancer patients will have a mutation in the *BRCA1 or BRCA2 gene*[20,21], who have high CBC risks. It has been shown that the relative risk associated with PRS is lower (for the first breast cancer) for women with a *BRCA1* and *BRCA2* mutation than in the general population[22], diluting the overall relative risk for CBC. More generally, it is possible that the CBC association may be attenuated due to the effect of other, unmeasured, genetic or other risk factors. If the risks are high, cases with higher PRS_{313} will have, on average, lower values of other risk factors, due to elimination of the highest risk individuals, again attenuating the CBC association. Finally, given the limited information on family history in our dataset, the estimate could have

been biased due to a family history effect not detected in our data.

There was some suggestion that the relative risk associated with PRS_{313} decreased with younger age, (P_{trend}=.05), and, specifically, was lower for women aged <40 years (HR per SD=1.13; 95%CI=0.98-1.31). Interestingly, Mavaddat et al[7] also found a lower relative risk below age 40 for first breast cancer. This effect may reflect the different characteristics of breast cancers at young ages, both in terms of germline susceptibility and pathology[23,24]. For example, the proportion of ER-negative breast cancers is higher at young ages, and the PRS is less predictive for ER-negative disease[6,7,24].

In the logistic regression analyses in Asian women, the association between the PRS_{313} and CBC risk was slightly weaker than in European women. This finding is consistent with a recent analysis investigating the association between a 287-variant PRS and first breast cancer risk in the Asian population[25], which showed an attenuated OR in Asian women (OR=1.52, 95%CI=1.49-1.56) compared to European women (OR=1.61, 95%CI=1.57-1.66). The lower estimate for Asian women might reflect the fact the PRS_{313} was developed in European populations, and the different LD structure in Asians may attenuate the association since the variants in the PRS are likely to be surrogates for the causal variants. Other explanations for the attenuated estimate may be the slightly younger age at first breast cancer diagnosis and the higher proportion ER-negative CBCs in Asian women compared to European women in our study. Finally, the imputation quality for variants was somewhat lower, on average, for the Asian than for the European dataset, with three variants on OncoArray and four variants on ICOGs with an imputation quality score<0.3 (Table S3). Nevertheless, we included those variants in the PRS for both European and Asian women, to keep the PRS comparable between ethnicities and studies. Future studies including larger numbers of Asian women, and women of other ethnicities, are needed to generate population-specific PRSs and to validate our findings in these groups.

A major strength of this study is the very large sample size in the BCAC dataset, including genotype information for ~150,000 women and a large number of CBC events. A limitation of this study is missing data on the patient, tumor, and treatment characteristics, which reduces the power of the multivariable Cox regression analyses and interaction analyses. In addition, registration of CBC was not complete; the 10-year cumulative CBC incidence was 2.2% in the BCAC dataset, compared to 3.8% using complete data from the Netherlands Cancer Registry[1]. For this reason, we estimated relative risk estimates using the BCAC data and applied these to external registry data to obtain absolute risk estimates. The underreporting of CBC should not bias our HR estimates, given that the event rate is low and reporting of CBC is unlikely to be related to the PRS_{313}. Moreover, we reran the cohort analysis in the subset of countries with a 10-year cumulative CBC incidence ≥3.0% in the BCAC dataset, and the estimates were very similar to the main analyses (HR per SD=1.23, 95%CI=1.14-1.33) (Figure S3).

In conclusion, the PRS_{313} is predictive for the development of CBC. We found no evidence for confounding or effect modification by other previously established CBC risk factors. The PRS_{313} is therefore likely to be an independent risk factor for CBC. Since the predictive ability of the PRS on its own is modest, it should be combined with other breast cancer risk factors to provide more useful CBC risk prediction models. More accurate risk prediction will help identify women at high CBC risk who will benefit from additional surveillance and/or risk reducing mastectomy, and equally important, to identify those women at low risk in order to avoid unnecessary surgeries.

Article information

Funding and acknowledgements

For full list of funding and acknowledgements see publication.

Conflict of interest

Dr. Beckmann conducts research funded by Amgen, Novartis and Pfizer, outside the submitted work. Dr. Fasching conducts research funded by Amgen, Novartis and Pfizer, outside the submitted work. He received honoraria from Roche, Novartis and Pfizer. Dr. Nevanlinna received honorarium from Astra Zeneca outside the submitted work.

Data and code availability

Data used in this manuscript may be requested through the original providers. Data of the Breast Cancer Association Consortium may be requested for non-profit research through an application procedure with the Breast Cancer Association Consortium; more information: http://bcac.ccge.medschl.cam.ac.uk/bcacdata/. Data of the UK Biobank needs to be requested through UK Biobank; more information: https://www.ukbiobank.ac.uk/researchers/

References

1 Kramer, I. *et al.* The influence of adjuvant systemic regimens on contralateral breast cancer risk and receptor subtype. *J. Natl. Cancer Inst.*, doi:10.1093/jnci/djz010 (2019).
2 Xiong, Z. *et al.* Patterns of Occurrence and Outcomes of Contralateral Breast Cancer: Analysis of SEER Data. *J Clin Med* **7**, doi:10.3390/jcm7060133 (2018).
3 Akdeniz, D. *et al.* Risk factors for metachronous contralateral breast cancer: A systematic review and meta-analysis. *Breast* **44**, 1-14, doi:10.1016/j.breast.2018.11.005 (2019).
4 Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92, doi:10.1038/nature24284 https://www.nature.com/articles/nature24284#supplementary-information (2017).
5 Milne, R. L. *et al.* Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat. Genet.* **49**, 1767-1778, doi:10.1038/ng.3785 (2017).
6 Mavaddat, N. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. *J. Natl. Cancer Inst.* **107**, doi:10.1093/jnci/djv036 (2015).
7 Mavaddat, N. *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* **104**, 21-34, doi:10.1016/j.ajhg.2018.11.002 (2019).
8 Brentnall, A. R. *et al.* A case-control evaluation of 143 single nucleotide polymorphisms for breast cancer risk stratification with classical factors and mammographic density. *Int. J. Cancer*, doi:10.1002/ijc.32541 (2019).
9 Shieh, Y. *et al.* Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res. Treat.* **159**, 513-525, doi:10.1007/s10549-016-3953-2 (2016).
10 Robson, M. E. *et al.* Association of Common Genetic Variants With Contralateral Breast Cancer Risk in the WECARE Study. *JNCI: Journal of the National Cancer Institute* **109**, djx051-djx051, doi:10.1093/jnci/djx051 (2017).
11 Teraoka, S. N. *et al.* Single nucleotide polymorphisms associated with risk for contralateral breast cancer in the Women's Environment, Cancer, and Radiation Epidemiology (WECARE) Study. *Breast Cancer Res.* **13**, R114, doi:10.1186/bcr3057 (2011).
12 Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat. Genet.* **47**, 373-380, doi:10.1038/ng.3242 (2015).
13 Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353-361, 361e351-352, doi:10.1038/ng.2563 (2013).
14 Amos, C. I. *et al.* The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol. Biomarkers Prev.* **26**, 126-135, doi:10.1158/1055-9965.epi-16-0106 (2017).
15 Schoenfeld, D. A. Sample-size formula for the proportional-hazards regression model. *Biometrics* **39**, 499-503 (1983).
16 Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *JAMA* **247**, 2543-2546 (1982).
17 Antoniou, A. C. *et al.* Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction. *Cancer Res.* **70**, 9742-9754, doi:10.1158/0008-5472.Can-10-1907 (2010).
18 Klevebring, D. *et al.* Exome sequencing of contralateral breast cancer identifies metastatic disease. *Breast Cancer Res. Treat.* **151**, 319-324, doi:10.1007/s10549-015-3403-6 (2015).
19 Begg, C. B. *et al.* Contralateral breast cancers: Independent cancers or metastases? *Int. J. Cancer* **142**, 347-356, doi:10.1002/ijc.31051 (2018).
20 Thompson, D. & Easton, D. The genetic epidemiology of breast cancer genes. *J. Mammary Gland Biol. Neoplasia* **9**, 221-236, doi:10.1023/B:JOMG.0000048770.90334.3b (2004).
21 van den Broek, A. J. *et al.* Impact of Age at Primary Breast Cancer on Contralateral Breast Cancer Risk in BRCA1/2 Mutation Carriers. *J. Clin. Oncol.* **34**, 409-418, doi:10.1200/jco.2015.62.3942 (2016).
22 Kuchenbaecker, K. B. *et al.* Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J. Natl. Cancer Inst.* **109**, doi:10.1093/jnci/djw302 (2017).
23 Azim, H. A., Jr. *et al.* Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. *Clin. Cancer Res.* **18**, 1341-1351, doi:10.1158/1078-0432.Ccr-11-2599 (2012).
24 Anders, C. K. *et al.* Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J. Clin. Oncol.* **26**, 3324-3330, doi:10.1200/jco.2007.14.2471 (2008).
25 Ho, W. K. *et al.* European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun* **11**, 3833, doi:10.1038/s41467-020-17680-w (2020).

Supplementary Material

Table S1. Study characteristics of included studies of the Breast Cancer Association Consortium

See online material

Table S2. Studies and samples included in the analyses using the case-case series, cohort, and validation set

	European								Asian		
	Case-case series N studies = 62			Cohort N studies = 42		Validation set N studies = 24			Case-case series N studies = 8		
Studies	Control women[a]	Unilateral BC	CBC	Unilateral BC	CBC	Control women[a]	Unilateral BC	CBC	Control women[a]	Unilateral BC	CBC
ABCFS	738	1,149	127	1,021	93	-	-	-	-	-	-
ABCS	1,567	1,047	54	519	14	-	-	-	-	-	-
ABCS-F	0	861	91	363	17	-	-	-	-	-	-
ABCTB	375	900	17	708	1	74	180	8	-	-	-
BBCC	711	845	58	766	6	49	56	5	-	-	-
BBCS	1,768	1,266	80	466	1	-	-	-	-	-	-
BCEES	-	-	-	-	-	166	133	0	-	-	-
BCFR-NY	27	340	61	-	-	-	-	-	-	-	-
BCFR-PA	0	104	14	69	4	-	-	-	-	-	-
BCFR-UTAH	0	13	87	-	-	-	-	-	-	-	-
BCINIS	-	-	-	-	-	144	262	0	-	-	-
BIGGS	49	713	50	395	2	-	-	-	-	-	-
BREOGAN	725	1,245	19	1,233	15	145	238	4	-	-	-
BSUCH	1,122	900	36	727	3	-	-	-	-	-	-
CBCS	817	530	21	-	-	163	105	4	170	238	10
CCGP	321	598	19	578	8	66	125	7	-	-	-
CGPS	5,250	4,135	60	3,834	17	142	227	3	-	-	-
CNIO-BCS	829	742	5	-	-	-	-	-	-	-	-
CTS	-	-	-	-	-	115	220	0	-	-	-
DIETCOMPLYF	0	704	1	-	-	-	-	-	-	-	-
FHRISK	0	119	2	-	-	-	-	-	-	-	-
GC-HBOC	1,732	2,690	230	1,406	47	-	-	-	-	-	-
GENICA	711	869	26	869	1	56	89	2	-	-	-
GESBC	181	303	3	-	-	-	-	-	-	-	-
GLACIER	0	1,733	230	-	-	-	-	-	-	-	-
HABCS	863	774	84	-	-	173	141	6	-	-	-
HCSC	0	362	13	273	9	-	-	-	-	-	-
HEBCS	1,060	1,632	116	1,578	41	-	-	-	-	-	-
HERPACC	-	-	-	-	-	-	-	-	1,659	756	18
HKBCS	-	-	-	-	-	-	-	-	451	403	12
HMBCS	345	729	28	-	-	-	-	-	-	-	-
HUBCS	116	198	2	-	-	-	-	-	-	-	-
ICICLE	1	138	12	-	-	-	-	-	-	-	-
KARBAC	0	761	46	443	32	-	-	-	-	-	-
KARMA	5,981	2,314	96	2,188	33	597	185	10	-	-	-
KBCP	431	516	9	-	-	-	-	-	-	-	-
KCONFAB/AOCS	898	397	83	305	26	-	-	-	-	-	-

Table S2. Continued

	European								Asian		
	Case-case series N studies = 62			Cohort N studies = 42		Validation set N studies = 24			Case-case series N studies = 8		
Studies	**Control women[a]**	**Unilateral BC**	**CBC**	**Unilateral BC**	**CBC**	**Control women[a]**	**Unilateral BC**	**CBC**	**Control women[a]**	**Unilateral BC**	**CBC**
LMBC	1,821	3,016	208	2,286	92	87	142	14	-	-	-
MABCS	88	80	9	74	2	-	-	-	-	-	-
MARIE	2,066	1,540	115	1,535	53	-	-	-	-	-	-
MBCSG	766	1,015	150	569 8					-	-	-
MCBCS	2,093	1,999	59	1,903	6	35	96	3	-	-	-
MCCS	1,207	1,034	2	-	-	142	86	0	-	-	-
MEC	1,123	1,016	38	988	23	-	-	-	-	-	-
MISS	1,529	582	6	563	3	304	83	0	-	-	-
MMHS	1,635	273	4	-	-	320	48	4	-	-	-
MYBRCA	-	-	-	-	-	-	-	-	4,197	3,652	105
NBCS	212	2,334	31	1,370	4	-	-	-	-	-	-
NBHS	-	-	-	-	-	122	79	0	-	-	-
NC-BCFR	150	614	69	602	5	-	-	-	52	391	33
NCBCS	1,006	1,988	42	-	-	-	-	-	-	-	-
OBCS	414	467	10	445	1	-	-	-	-	-	-
OFBCR	728	1,908	143	1,656	51	-	-	-	-	-	-
ORIGO	0	1,090	89	1,053	69	132	134	15	-	-	-
PBCS	2,082	1,719	40	1,625	9	331	215	2	-	-	-
PKARMA	5,435	4,81	277	4,685	124	1	4	0	-	-	-
POSH	0	1,069	19	1,063	16	-	-	-	-	-	-
PREFACE	0	2,73	90	-	-	-	-	-	-	-	-
PROCAS	1,647	488	9	422	3	-	-	-	-	-	-
RBCS	0	873	152	724	81	-	-	-	-	-	-
SASBAC	1,378	1,118	22	1,086	5	-	-	-	-	-	-
SBCS	848	748	14	691	1	-	-	-	-	-	-
SEARCH	9,056	12,423	118	12,117	59	197	628	0	-	-	-
SEBCS	-	-	-	-	-	-	-	-	2,236	2,080	21
SGBCC	-	-	-	-	-	-	-	-	4,141	1,250	124
SKKDKFZS	29	1,084	71	1,054	41	-	-	-	-	-	-
SMC	-	-	-	-	-	141	244	0	-	-	-
SUCCESSB	0	438	2	-	-	-	-	-	-	-	-
SUCCESSC	0	2,807	29	-	-	-	-	-	-	-	-
SZBCS	489	676	6	409	1	-	-	-	-	-	-
TNBCC	152	1,037	2	-	-	-	-	-	-	-	-
TWBCS	-	-	-	-	-	-	-	-	492	1,250	17
UCIBCS	258	397	1	380	1	51	61	7	-	-	-

Table S2. Continued

	European								Asian		
	Case-case series N studies = 62			Cohort N studies = 42		Validation set N studies = 24			Case-case series N studies = 8		
Studies	**Control women[a]**	**Unilateral BC**	**CBC**	**Unilateral BC**	**CBC**	**Control women[a]**	**Unilateral BC**	**CBC**	**Control women[a]**	**Unilateral BC**	**CBC**
Total	**62,830**	**81,000**	**3,607**	**55,041**	**1,027**	**3,753**	**3,781**	**94**	**13,398**	**12,133**	**340**
Characteristics											
Invasiveness in situ	-	excluded	361	excluded	104	-	3[b]	7	-	excluded	67
invasive	-	79,876	2,200	54,675	670	-	3,777	60	-	11,929	209
unknown	-	1,124	1,046	366	253	-	1	27	-	204	64
ER status negative	-	13,828	446	9,333	105	-	766	8	-	3,457	54
positive	-	52,238	2,048	37,420	289	-	3,001	47	-	7,826	163
unknown	-	14,934	1,113	8,288	633	-	14	39	-	850	123

Abbreviations: BC = breast cancer; CBC = contralateral breast cancer; ER = estrogen receptor

[a] Without any diagnosis of breast cancer

[b] Due to the use of a new freeze of the BCAC data, N=3 breast cancers were now defined as in situ, which had previously been defined as invasive; the original validation dataset contained data of two additional studies[1]

Table S3. Variant information and breast cancer risk coefficients for the 77-variant PRS, 313-variant PRS, and ER-specific PRSs; previously published in Mavaddat et al.[1,2]

See online material

Table S4. Patient, tumor, and treatment characteristics of all women diagnosed with first invasive breast cancer since 1990 (European cohort)

Characteristics		Number of women (%)[a]
Total		56,068 (100)
Median age at first diagnosis in years (range)		56 (18-98)
Year of diagnosis		
	1990-1994	3,029 (5.4)
	1995-1999	10,153 (18.1)
	2000-2004	18,484 (33.0)
	2005-2009	17,575 (31.3)
	2010-2015	6,827 (12.2)
Family history (first degree relative)		
	no	33,623 (76.4)
	yes	10,369 (23.6)
	unknown	12,076
Nodal status		
	negative	29,070 (61.9)
	positive	17,903 (38.1)
	unknown	9,095
Tumor size, cm		
	≤2	28,057 (63.8)
	(2, 5]	14,138 (32.2)
	>5	1,750 (4.0)
	unknown	12,123
Differentiation grade		
	I	8,721 (19.5)
	II	21,621 (48.3)
	III	14,454 (32.3)
	unknown	11,272
Morphology		
	ductal	37,324 (76.6)
	lobular	5,878 (12.1)
	mixed (ductal and lobular)	2,174 (4.5)
	other	3,344 (6.9)
	unknown	7,348
ER-status		
	negative	9,527 (20.0)
	positive	38,090 (80.0)
	unknown	8,451
PR-status		
	negative	13,098 (32.6)
	positive	27,044 (67.4)
	unknown	15,926
HER2-status		
	negative	23,787 (82.7)
	positive	4,969 (17.3)
	unknown	27,312
Surgery		
	yes, breast saving	16,468 (42.3)
	yes, mastectomy	11,315 (29.1)
	yes, type unknown	11,163 (28.7)
	unknown	17,122
(Neo)adjuvant chemotherapy		
	no	18,110 (49.4)
	yes	18,559 (50.6)
	unknown	19,399
(Neo)adjuvant endocrine therapy		
	no	10,781 (28.3)
	yes	27,322 (71.7)
	unknown	17,965
Radiotherapy		
	no	11,023 (27.4)
	yes	29,142 (72.6)
	unknown	15,903

Abbreviations: ER = estrogen receptor; PR = progesterone receptor; HER2 = human epidermal growth factor receptor 2

[a] Total may not be 100% because of rounding

Table S5. Association between the 313-variant PRS (PRS$_{313}$) and contralateral breast cancer risk in the European cohort

Percentile categories of the PRS$_{313}$	No. of women	No. of CBC	HR per unit SD[a]	95%CI	P-value
0th to 10th	5,607	65	0.59	0.45-0.78	<.001
10th to 20th	5,606	79	0.71	0.55-0.92	.01
20th to 40th	11,214	165	0.74	0.60-0.90	.003
40th to 60th	11,214	224	1.00	Ref.	-
60th to 80th	11,214	208	0.90	0.74-1.08	.25
80th to 90th	5,607	121	1.05	0.84-1.31	.69
90th to 100th	5,606	165	1.38	1.13-1.69	.002

Abbreviations: PRS = polygenic risk score; No = number; CBC = contralateral breast cancer; HR = hazard ratio; CI = confidence interval; SD = standard deviation

[a]The analysis was performed with attained age as time scale. Coefficients to construct the PRS$_{313}$ are shown in Table S3. The PRS$_{313}$ was standardized by SD=0.61, in line with Mavaddat et al.[1]

Table S6. Multivariable Cox regression models of contralateral breast cancer risk by 313-variant PRS (PRS$_{313}$) in all women, all women excluding studies oversampling cases with family history, and those with complete covariate information

		All patients			All women excluding studies oversampling cases with family history			Complete case		
		N=56,068 (CBC=1,027)			N=51,883 (CBC=829)			N=12,065 (CBC=193)		
		HR per unit SD[a]	95%CI	P-value	HR per unit SD[a]	95%CI	P-value	HR per unit SD[a]	95%CI	P-value
Model 1										
PRS$_{313}$[b]		1.25	1.18-1.33	<.001	1.26	1.17-1.34	<.001	1.35	1.17-1.56	<.001
Model 2										
PRS$_{313}$[b]		1.23	1.16-1.31	<.001	1.25	1.17-1.34	<.001	1.33	1.15-1.54	<.001
Family history	yes vs. no	1.43	1.24-1.64	<.001	1.34	1.13-1.59	.001	1.49	1.06-2.09	.02
	unknown vs. no	0.93	0.75-0.16	.54	0.92	0.73-1.16	.47	-	-	-
Model 3										
PRS$_{313}$[b]		1.25	1.18-1.33	<.001	1.26	1.17-1.34	<.001	1.35	1.17-1.56	<.001
Nodal status	positive vs. negative	1.05	0.91-1.20	.50	1.07	0.92-1.25	.37	1.14	0.85-1.53	.37
	unknown vs. no	1.26	1.04-1.53	.02	1.29	1.04-1.60	.02	-	-	-
Model 4										
PRS$_{313}$[b]		1.25	1.18-1.33	<.001	1.26	1.18-1.35	<.001	1.35	1.17-1.56	<.001
Tumor size,	(2-5] vs. ≤2	1.08	0.92-1.25	.34	1.12	0.95-1.32	.20	0.93	0.68-1.27	.66
	>5 vs. ≤2	1.37	0.99-1.89	.06	1.45	1.02-2.07	.04	1.63	0.93-2.85	.09
	unknown vs. ≤2	1.23	1.04-1.47	.02	1.14	0.94-1.39	.18	-	-	-
Model 5										
PRS$_{313}$[b]		1.25	1.17-1.33	<.001	1.25	1.17-1.34	<.001	1.35	1.17-1.57	<.001
Differentiation grade	II vs. I	0.93	0.76-1.13	.45	0.99	0.80-1.24	.94	0.98	0.65-1.48	.93
	III vs. I	0.90	0.73-1.12	.35	0.97	0.76-1.24	.81	1.09	0.70-1.69	.69
	unknown vs. I	1.20	0.96-1.49	.11	1.45	1.13-1.86	.004	-	-	-
Model 6										
PRS$_{313}$[b]		1.25	1.17-1.33	<.001	1.25	1.17-1.34	<.001	1.33	1.16-1.54	<.001
Morphology	lobular vs. ductal	1.26	1.03-1.53	.03	1.34	1.08-1.67	.008	1.48	0.99-2.21	.05
	mixed (ductal and lobular) vs. ductal	1.28	0.94-1.73	.11	1.36	0.98-1.88	.06	1.48	0.87-2.54	.15
	other vs. ductal	1.04	0.81-1.33	.75	0.91	0.66-1.24	.55	1.24	0.69-2.21	.47
	unknown vs. ductal	1.77	1.42-2.19	<.001	1.82	1.44-2.30	<.001	-	-	-
Model 7										
PRS$_{313}$[b]		1.25	1.18-1.33	<.001	1.26	1.18-1.35	<.001	1.35	1.17-1.56	<.001
ER-status	positive vs. negative	0.88	0.75-1.04	.14	0.86	0.72-1.03	.11	0.90	0.62-1.32	.60
	unknown vs. negative	1.16	0.93-0.43	.19	1.11	0.86-1.43	.43	-	-	-
Model 7										
PRS$_{313}$[b]		1.25	1.18-1.33	<.001	1.26	1.18-1.35	<.001	1.35	1.17-1.56	<.001
PR-status	positive vs. negative	0.95	0.81-1.11	.51	0.92	0.78-1.09	.32	0.91	0.66-1.25	.56
	unknown vs. negative	1.15	0.95-1.40	.14	1.10	0.88-1.37	.40	-	-	-

Table S6. Continued

		All patients			All women excluding studies oversampling cases with family history			Complete case		
		N=56,068 (CBC=1,027)			N=51,883 (CBC=829)			N=12,065 (CBC=193)		
		HR per unit SD[a]	95%CI	P-value	HR per unit SD[a]	95%CI	P-value	HR per unit SD[a]	95%CI	P-value
Model 9										
PRS_{313}[b]		1.25	1.18-1.33	<.001	1.26	1.17-1.34	<.001	1.34	1.16-1.55	<.001
HER2-status	positive vs. negative	0.84	0.64-1.11	.22	0.76	0.56-1.05	.10	0.70	0.45-1.10	.12
	unknown vs. negative	1.29	1.11-1.50	.001	1.28	1.08-1.52	.004	-	-	-
Model 10										
PRS_{313}[b]		1.25	1.17-1.32	<.001	1.26	1.17-1.34	<.001	1.35	1.16-1.56	<.001
Chemotherapy	yes vs. no	0.86	0.73-1.01	.06	0.99	0.83-1.19	.92	0.89	0.64-1.25	.51
	unknown vs. no	1.09	0.91-1.31	.34	1.20	0.97-1.47	.09	-	-	-
Model 11										
PRS_{313}[b]		1.25	1.18-1.33	<.001	1.26	1.18-1.35	<.001	1.36	1.17-1.57	<.001
Endocrine therapy	yes vs. no	0.75	0.64-0.88	.001	0.92	0.75-1.12	.41	0.78	0.55-1.11	.17
	unknown vs. no	0.90	0.75-1.09	.28	1.11	0.87-1.41	.39	-	-	-
Model 12										
PRS_{313}[b]		1.25	1.17-1.32	<.001	1.26	1.17-1.34	<.001	1.35	1.17-1.56	<.001
Radiotherapy	yes vs. no	1.00	0.85-1.18	1.00	0.98	0.82-1.18	.85	1.35	0.88-2.08	.17
	unknown vs. no	1.41	1.14-1.74	.001	1.18	0.93-1.50	.17	-	-	-
Model 13										
PRS_{313}[b]		1.25	1.17-1.32	<.001	1.25	1.17-1.34	<.001	1.34	1.16-1.55	<.001
Year of first breast cancer diagnosis		0.95	0.94-0.96	<.001	0.95	0.93-0.96	<.001	0.90	0.86-0.95	<.001
Model 14										
PRS_{313}[b]	full model[c]	1.23	1.16-1.31	<.001	1.25	1.16-1.33	<.001	1.33	1.15-1.53	<.001

Abbreviations: PRS = polygenic risk score; CBC = contralateral breast cancer; HR = hazard ratio; CI = confidence interval; SD = standard deviation; ER = estrogen receptor; PR = progesterone receptor; HER2 = human epidermal growth factor receptor 2

[a] All analyses were performed with attained age as the time scale

[b] Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al.[1]

[c] Adjusted for family history, nodal status, tumor size, differentiation grade, morphology, ER status, HER2 status, chemotherapy, endocrine therapy, radiotherapy, and year of first breast cancer diagnosis

Table S7. Five- and ten-year cumulative risks of contralateral breast cancer by the 313-variant PRS (PRS_{313}) for different age groups

	5-year cumulative CBC risks (%) range by age					10-year cumulative CBC risks (%) range by age				
Age at first breast cancer diagnosis (years)	**5th percentile PRS_{313}**	**10th percentile PRS_{313}**	**50th percentile PRS_{313}**	**90th percentile PRS_{313}**	**95th percentile PRS_{313}**	**5th percentile PRS_{313}**	**10th percentile PRS_{313}**	**50th percentile PRS_{313}**	**90th percentile PRS_{313}**	**95th percentile PRS_{313}**
30-34	1.9-3.3	2.1-3.6	2.8-4.7	3.7-6.2	4.0-6.8	3.3-4.4	3.6-4.8	4.7-6.3	6.2-8.3	6.8-9.1
35-39	0.8-2.2	0.9-2.4	1.2-3.2	1.6-4.2	1.7-4.6	2.2-3.9	2.4-4.2	3.2-5.5	4.2-7.2	4.6-8.0
40-44	1.5-2.9	1.7-3.2	2.2-4.2	2.9-5.5	3.2-6.0	2.9-4.9	3.2-5.3	4.2-7.0	5.5-9.1	6.0-10.0
45-49	1.4-2.5	1.5-2.8	2.0-3.7	2.6-4.8	2.9-5.3	2.5-4.2	2.8-4.5	3.7-6.0	4.8-7.8	5.3-8.6
50-54	1.4-2.9	1.5-3.1	2.0-4.1	2.6-5.5	2.9-6.0	2.9-4.8	3.1-5.3	4.1-6.9	5.5-9.1	6.0-10.0
55-59	1.6-3.3	1.8-3.6	2.4-4.7	3.1-6.2	3.4-6.8	3.3-5.3	3.6-5.7	4.7-7.5	6.2-9.8	6.8-10.8
60-64	1.8-3.5	1.9-3.8	2.6-5.0	3.4-6.5	3.7-7.2	3.5-5.5	3.8-6.0	5.0-7.9	6.5-10.3	7.2-11.3
65-70	1.5-3.5	1.7-3.8	2.2-5.0	2.9-6.6	3.2-7.2	3.5-4.6	3.8-5.0	5.0-6.6	6.6-8.7	7.2-9.5

Abbreviations: PRS = polygenic risk score; CBC = contralateral breast cancer

Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al[1]. The CBC incidences for each age group were calculated based on incidence data from the Netherlands Cancer Registry[3] and relative risks estimated as described in the Material and Methods. In contrast to Table 4, death was not taken into account as competing risk

Table S8. Estimates of unilateral- and contralateral breast cancer risk by the 313-variant PRS (PRS_{313}) in the European case-case series and the Asian case-case series

	European						Asian		
	Case-case series[a]			Validation set[b]			Case-case series[a]		
PRS_{313}[c]	**OR per unit SD**	**95%CI**	**P-value**	**OR per unit SD**	**95%CI**	**P-value**	**OR per unit SD**	**95%CI**	**P-value**
Unilateral breast cancer versus control	1.82	1.80-1.84	<.001	1.67	1.59-1.76	<.001	1.56	1.52-1.60	<.001
CBC versus unilateral breast cancer	1.30	1.26-1.35	<.001	1.39	1.13-1.70	.002	1.15	1.02-1.29	.02

Abbreviations: PRS = polygenic risk score; CBC = contralateral breast cancer; OR = odds ratio; SD = standard deviation; CI = confidence interval

[a] Adjusted for country and age. For all women with unilateral- and contralateral breast cancer we used age at first breast cancer diagnosis, and for control women without any diagnosis of breast cancer we used age at baseline questionnaire

[b] The validation set was previously used to develop the PRS_{313}; see details in materials and methods. For analyses in the current paper, this set is nested within the case-case series. These analyses were additionally adjusted for 10 principal components for comparability with the originally published PRS_{313} overall estimates[1]

[c] Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al.[1]

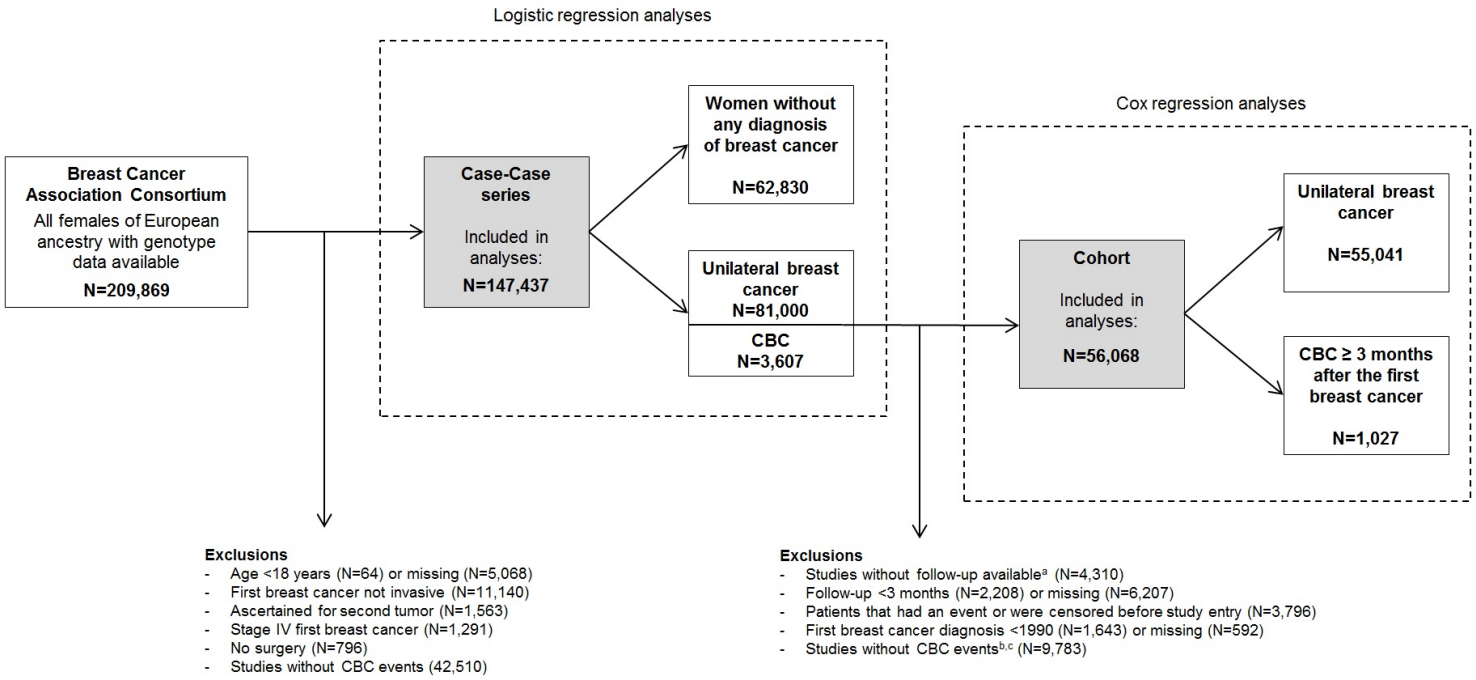

Figure S1A. Overview of the selection of women with breast cancer and control women for the European series
Abbreviations: CBC = contralateral breast cancer. For a complete overview of all studies see Table S1. [a] Excluded studies: CBCS, GLACIER, HMBCS, TNBCC; [b] Excluded studies: BCFR-NY, BCFR-UTAH, CNIO-BCS, DIETCOMPLYF, FHRISK, GESBC, HABCS, HUBCS, ICICLE, KBCP, MCCS, MMHS, NCBCS, PREFACE, SUCCESSB, SUCCESSC; [c] These studies dropped out because for these analyses the definition of CBC is based on the criteria that the CBC was diagnosed at least three months after the first breast cancer diagnosis

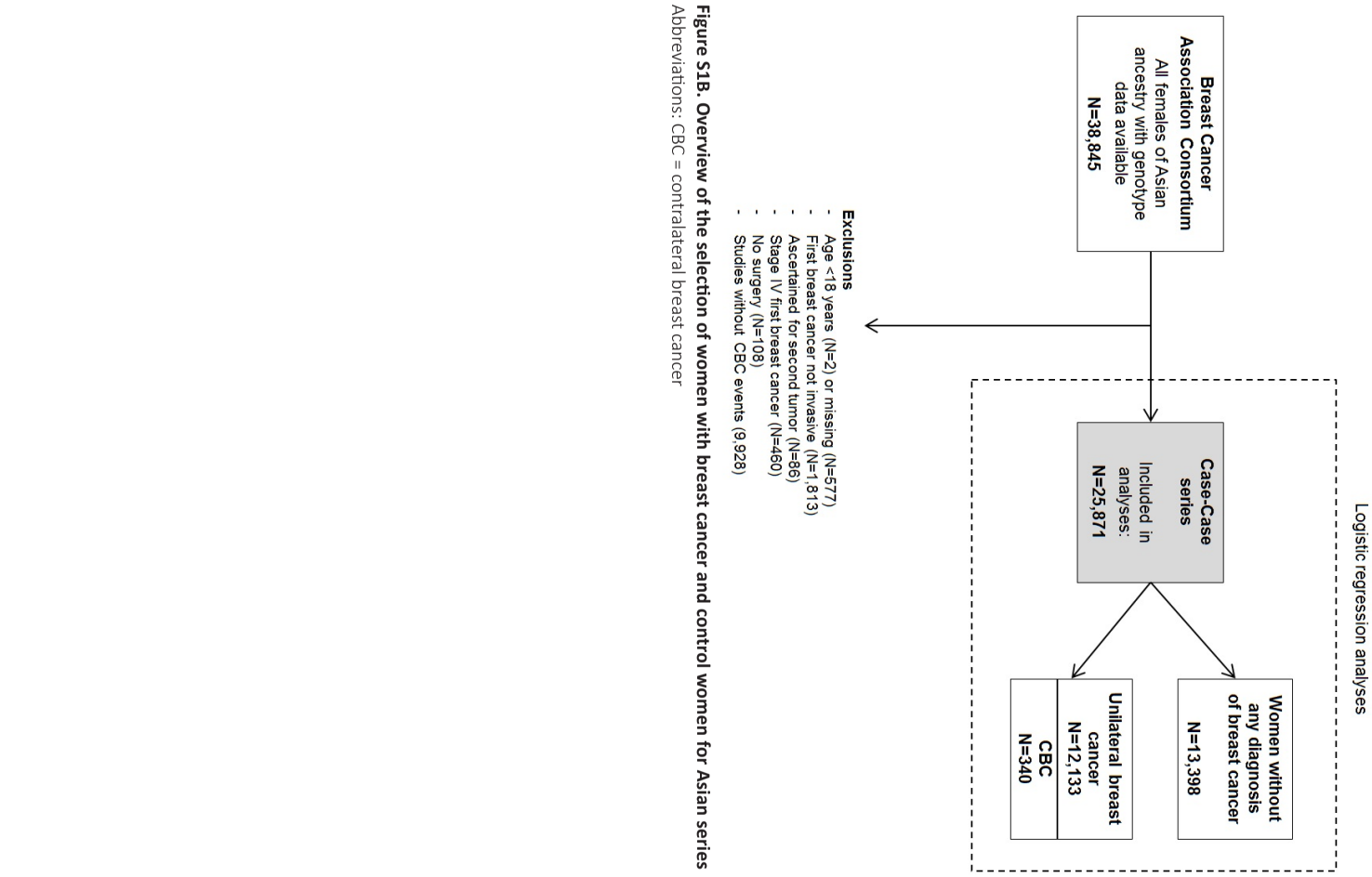

Figure S1B. Overview of the selection of women with breast cancer and control women for Asian series
Abbreviations: CBC = contralateral breast cancer

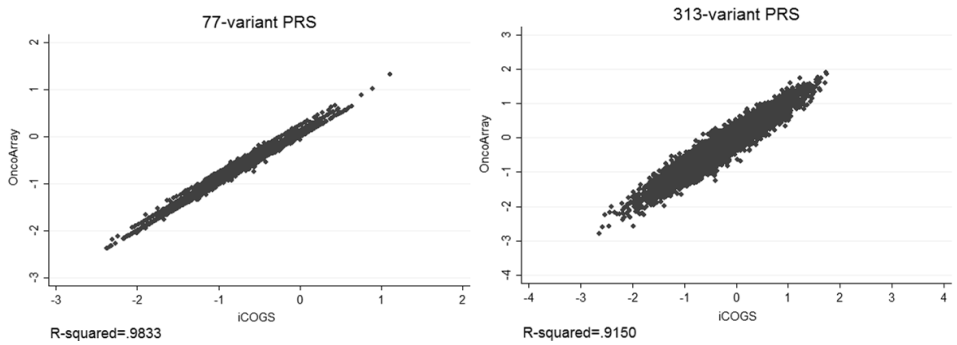

Figure S2. Correlation of total variant scores between the iCOGS array and OncoArray for the 77-variant PRS and the 313-variant PRS[a,b]

Abbreviations: PRS = polygenic risk score; SD = standard deviation. [a] We evaluated consistency between iCOGS and OncoArray using the intraclass correlation coefficient (ICC), showing a ICC of 0.99 (95%CI=0.99-0.99) for the PRS_{77}, and an ICC of 0.96 (95%CI=0.95-0.96) for the PRS_{313}, based on N=9,071 observations; [b] Coefficients to construct the PRSs are shown in Table S3. The PRSs were standardized by the same SD as was used by Mavaddat et al.[1]. The SD was 0.45 for overall breast cancer PRS_{77}, and 0.61 for overall breast cancer PRS_{313}

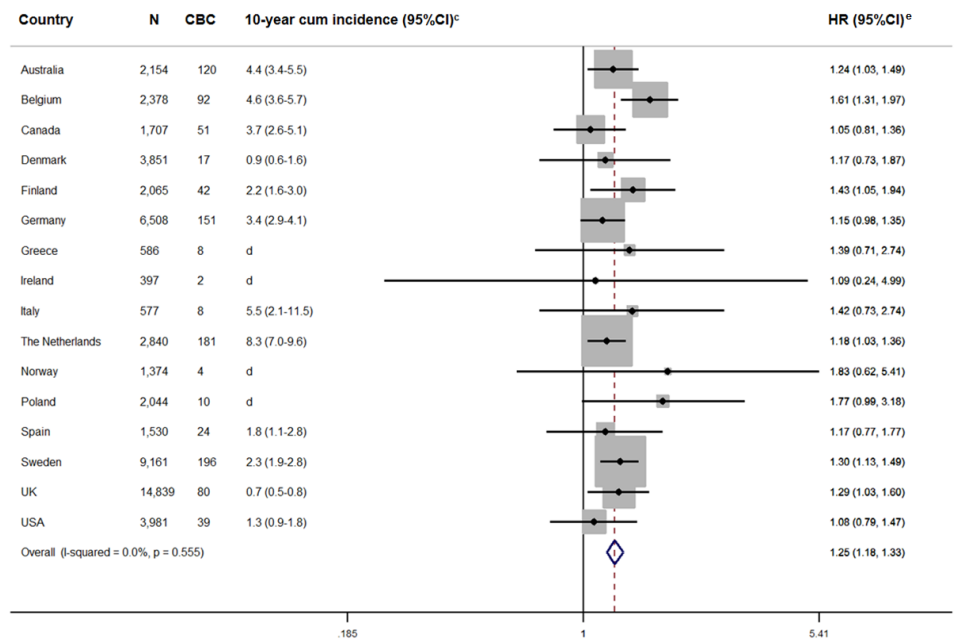

Figure S3. Forest plot of the association between the 313-variant PRS and contralateral breast cancer risk by country[a,b]

Abbreviations: PRS = polygenic risk score; N = number of women; CBC = contralateral breast cancer; cum = cumulative; CI = confidence interval; HR = hazard ratio; SD = standard deviation. Fixed effect meta-analysis was used to calculate I-squared and P-value for heterogeneity. [a] Republic of North Macedonia was left out this plot because of a too small sample size (N=76 women including N=2 CBC events); [b] Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al.[1]; [c] The 10-year cumulative incidence of CBC was estimated with time since first breast cancer as time scale, and distant metastases (where available) and death as competing risks; [d] Follow-up too short for calculating 10-year cumulative incidence; [e] HR per SD. The analyses were performed with attained age as the time scale

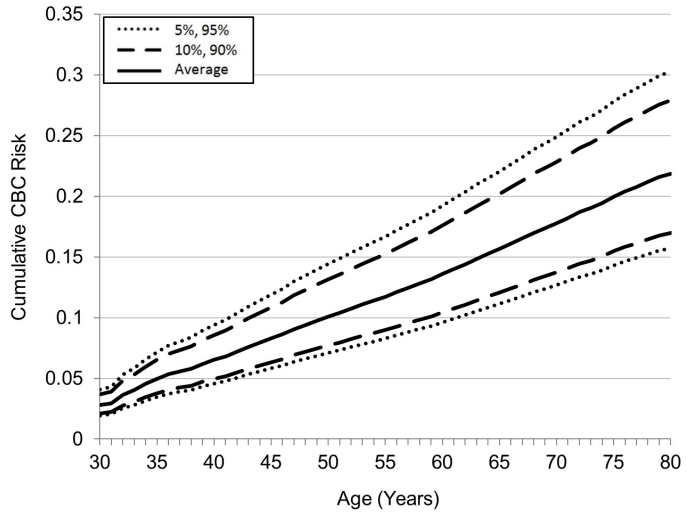

Figure S4. Predicted contralateral breast cancer risk by percentile of the 313-variant PRS (PRS_{313})
Abbreviations: PRS = polygenic risk score, CBC = contralateral breast cancer
Coefficients to construct the PRS_{313} are shown in Table S3. The PRS_{313} was standardized by SD=0.61, in line with Mavaddat et al[1]. The CBC incidences were calculated based on incidence data from the Netherlands Cancer Registry[3] and relative risks estimated as described in the Material and Methods. In contrast to Figure 2, death was not taken into account as competing risk

Supplemental Note

Our initial aim was to externally validate our results using the UK Biobank, which seemed the most suitable cohort given the large number of women diagnosed with breast cancer with information available on the PRS_{313},. However, when we started the analyses, it turned out that the UK Biobank had no information available on the laterality of the second breast tumor. Therefore, we were unable to distinguish between ipsilateral and contralateral breast cancer, and had to define our endpoint in these analyses as 'any second breast cancer'. In addition, in comparison to our analyses in the BCAC, we were unable to exclude patients diagnosed with stage IV invasive first breast cancer from the UK Biobank cohort, and had limited information on metastases developed during follow-up.

The association between the overall breast cancer PRS_{313} and (any) second breast cancer was evaluated among women aged ≥18 years of European ancestry from the UK Biobank cohort who had had a diagnosis of invasive first breast cancer. UK Biobank samples were genotyped using Affymetrix UK BiLEVE Axiom array and Affymetrix UK Biobank Axiom® array and imputed to the combined 1000 Genome Project v3 and UK10K reference panels using SHAPEIT3 and IMPUTE3[4]. The lowest imputation info score for the variants used in these analyses was 0.86. Samples were included for this analysis of the UK BIOBANK study on the basis of female sex (genetic and self-reported) and ethnicity filter (Europeans/White British ancestry subset). Duplicates and individuals with high degree of relatedness (samples which have >10 putative third degree relatives) were removed, and we randomly excluded one of each related pair first-degree relatives. Samples were also excluded on standard quality control criteria. The PRS_{313} was calculated as a weighted sum of the minor allele dosages; the variant selection and weights are as given by Mavaddat et al[1]. The PRS_{313} was standardized by SD=0.61, in line with our BCAC analyses and Mavaddat et al[1].

The final cohort included 10,567 women with invasive breast cancer among whom 302 registry-confirmed second breast cancers developed over 59,260 person-years of follow-up. A Cox proportional hazards model was used to assess the association between PRS_{313} and second breast cancer risk. Time at risk started three months after the age of first breast cancer diagnosis, where this was diagnosed after the baseline questionnaire date, or three months after the baseline questionnaire where first breast cancer was diagnosed before the baseline questionnaire date. Time at risk ended at the age of second breast cancer diagnosis (ipsilateral or contralateral), distant metastasis (where available), death or end of follow-up (at latest December 10, 2016). Potential effect modification of the PRS_{313} by age was evaluated by adding an interaction term (PRS_{313} x age at first breast cancer diagnosis [continuous]) in the model. We performed a separate analysis for invasive second breast cancer (241 breast cancers), where we censored on in situ second breast cancer.

The HR for a second breast cancer (in situ or invasive) per SD of PRS_{313} in the UK Biobank cohort was 1.13 (95%CI=1.01-1.26). We found no indication for interaction with age at first breast cancer diagnosis ($HR_{interaction}$=1.00, 95%CI=0.99-1.01; P=0.87). When analyses were restricted to invasive second breast cancer, the HR per SD was 1.13 (95%CI=1.00-1.29).

Supplemental References

1 Mavaddat, N. *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* **104**, 21-34, doi:10.1016/j.ajhg.2018.11.002 (2019).

2 Mavaddat, N. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. *J. Natl. Cancer Inst.* **107**, doi:10.1093/jnci/djv036 (2015).

3 Kramer, I. *et al.* The influence of adjuvant systemic regimens on contralateral breast cancer risk and receptor subtype. *J. Natl. Cancer Inst.*, doi:10.1093/jnci/djz010 (2019).

4 O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nat. Genet.* **48**, 817-820, doi:10.1038/ng.3583 (2016).