The path to individualised breast cancer screening
Lakeman, I.M.M.

Citation
Lakeman, I. M. M. (2022, June 14). *The path to individualised breast cancer screening*. Retrieved from https://hdl.handle.net/1887/3420638

Version: Publisher's Version

CHAPTER 2

2

Addition of a 161-SNP Polygenic Risk Score to family history-based risk prediction: impact on clinical management in non-*BRCA1/2* breast cancer families

Inge M.M. Lakeman[#], Florentine S. Hilbers[#], Mar Rodriguez-Girondo, Andrew Lee, Maaike P.G. Vreeswijk, Antoinette Hollestelle, Caroline Seynaeve, Hanne E.J. Meijers-Heijboer, Jan C. Oosterwijk, Nicoline Hoogerbrugge, Edith Oláh, Hans F.A. Vasen, Christi J. van Asperen, and Peter Devilee

[#]these authors contributed equally

Lakeman IMM, et al. Journal of Medical Genetics 2019;56:581-589

Abstract

Background

The currently known breast cancer associated Single Nucleotide Polymorphisms (SNPs) are presently not used to guide clinical management. We explored whether a genetic test that incorporates a SNP-based Polygenic Risk Score (PRS) is clinically meaningful in non-*BRCA1/2* high-risk breast cancer families.

Methods

101 non-BRCA1/2 high-risk breast cancer families were included; 323 cases and 262 unaffected female relatives were genotyped. The 161-SNP PRS was calculated and standardised to 327 population controls (sPRS). Association analysis was performed using a Cox-type random effect regression model adjusted by family history. Updated individualised breast cancer lifetime risk scores were derived by combining the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) breast cancer lifetime risk with the effect of the sPRS.

Results

The mean sPRS for cases and their unaffected relatives was 0.70 (SD=0.9) and 0.53 (SD=0.9), respectively. A significant association was found between sPRS and breast cancer, HR=1.16, 95%CI=1.03-1.28, p=0.026. Addition of the sPRS to risk prediction based on family history alone changed screening recommendations in 11.5%, 14.7%, and 19.8% of the women according to breast screening guidelines from the USA (National Comprehensive Cancer Network), UK (National Institute for Health and Care Excellence and the Netherlands (Netherlands Comprehensive Cancer Organisation), respectively.

Conclusion

Our results support the application of the PRS in risk prediction and clinical management of women from genetically unexplained breast cancer families.

Introduction

Breast cancer is the most common cancer in women in the Western world. For women with a first-degree relative with breast cancer, the risk for developing breast cancer is twofold in comparison with women without such a family history[1]. Approximately 20% of this familial relative risk is explained by pathogenic variants in the high-risk genes *BRCA1* and *BRCA2*, 2-5% by variants in other breast cancer genes (e.g. *CHEK2*, *PALB2*, and *ATM*), and 18% by the currently known common low risk variants, mostly single nucleotide polymorphisms (SNPs)[2-5].

Individually these SNPs confer a very small increase in breast cancer risk but jointly they may confer a substantial increase of the risk[2]. This combined risk of all SNPs associated with breast cancer can be summarised in a Polygenic Risk Score (PRS). The PRS can stratify women into different risk categories[2, 6-8], which for 8% of women from the general population might be high enough to be clinically relevant, regardless of family history[2].

The PRS may also be combined with other risk factors, such as *BRCA1/2* status or breast cancer family history, to further refine and individualise risk estimation. The large majority of breast cancer families seen in Family Cancer Clinics today cannot be linked to pathogenic variants in *BRCA1* or *BRCA2*. Risk management for women from these families is based mainly on family history, which can be used as a variable to calculate individual breast cancer risk in various risk prediction algorithms[9], such as the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA)[10].

Until now, the PRS is not included in clinical genetic practice to guide clinical management. Several studies have shown an improved discriminative power between breast cancer cases and controls by combining the PRS with a breast cancer risk prediction tool[11-14]. However, little is currently known of the discriminative power of the PRS between family members, with respect to who will develop breast cancer. A recent study genotyped cases and controls in 52 Finnish non-*BRCA1/2* breast cancer families to calculate a 75-SNP PRS. The PRS for healthy women from breast cancer families was lower in comparison to affected family members[15]. This suggests that the PRS can help to individualise risk stratification and advice for surveillance for women in breast cancer families.

Here, we explore the clinical applicability of the 161-SNP PRS for risk prediction in a cohort of 101 high-risk breast cancer families not explained by pathogenic variants in the *BRCA1* and *BRCA2* genes. The clinical impact of the PRS on breast cancer risk prediction based on family history alone was investigated by determining the potential change in clinical management, as stipulated by three currently used guidelines (the National Comprehensive Cancer Network guideline (NCCN)[16], the National Institute for Health and Care Excellence guideline (NICE)[17], and the Netherlands Comprehensive Cancer Organisation guideline (IKNL)[18]).

Materials and Methods

Study cohorts

Two cohorts were included, a hospital-based case-control (Oorsprong van borstkanker integraal onderzocht (ORIGO)) and a family-based case-control cohort. Informed consent was obtained for all individuals. Population controls were irreversibly anonymised. Only women were included in this study.

The ORIGO cohort consists of incident breast cancer cases, not selected for breast cancer family history enrolled between 1996 and 2006 in the context of the ORIGO study, as described elsewhere[19]. For the present study, 357 ORIGO cases were selected for which genotyping had been performed on the iCOGS array. Likewise, 327 healthy genotyped bloodbank donors were included in the ORIGO cohort as controls. Age of last follow up was determined as the age at diagnosis for cases and the age at inclusion for controls.

The families from the family-based cohort were selected between 1990 and 2012 through five Clinical Genetic Services (Rotterdam, Groningen, Nijmegen, Leiden, the Netherlands and Budapest, Hungary) and the Foundation for the Detection of Hereditary Tumours in the Netherlands, as previously described[20]. At least one family member affected with breast cancer was tested for *BRCA1* and *BRCA2*. We did not have informed consent for testing other specific genes besides *BRCA1* and *BRCA2*. The selection criteria for families included: breast cancer (invasive/in situ) before the age of 60 in at least three women, or in two women if at least one of them had bilateral breast cancer before the age of 60. In total, 102 families without a pathogenic variant in *BRCA1* or *BRCA2* were included of which a blood DNA sample was available for 612 women. Of these women, 340 were affected with breast cancer and 272 were unaffected relatives. The unaffected relatives were censored regarding breast cancer, irrespective of other types of cancer. Most cancers were verified with a pathology report. Date of last follow up was determined as the date of last contact with the family.

Genotyping

DNA samples of all included individuals were genotyped with the iCOGS SNP array, designed for association analysis in breast, ovarian and prostate cancer, containing 211,155 SNPs[3]. Genotyping and quality control of the ORIGO cohort was performed as part of association studies conducted by the Breast Cancer Association Consortium (BCAC)[3]. For the family-based cohort, quality control led to the exclusion of 27 individuals (see supplementary material and methods). Therefore further analysis was done with 323 breast cancer cases and 262 unaffected relatives from 101 families for this cohort.

Imputation

Some of the 182 currently known SNPs are associated primarily with Estrogen Receptor (ER)-negative or ER-positive breast cancer. We constructed a PRS for overall breast cancer with 161 SNPs, selecting all SNPs significantly associated ($p<5.10^{-8}$) with overall breast cancer in case-control studies performed by BCAC[4] (Table S1). ER-status was not known for all cases in our study, and substrata would become too small to reach sufficient statistical power for ER-specific PRSs. The 85 SNPs that were not directly genotyped by the iCOGS array were imputed by pre-phasing with SHAPEIT and IMPUTE2[21, 22]. To improve imputation quality both the reference panels 1000 genomes phase 3 and GoNL were used[23, 24].

Polygenic risk score

The following formula was used to calculate the PRS based on 161 SNPs:

$$PRS_j = \sum_{i=1}^{161} n_{ij} \ln{(OR_i)}$$

where n_{ij} is the number of risk alleles (0, 1 or 2) for SNP i carried by individual j and OR_i is the per-allele log odds ratio (OR) for breast cancer associated with SNP i. The ORs were the most recent estimates from analysis of the OncoArray data[4] (Table S1). The majority of studies used for this analysis were population-based case–control studies[4].

The PRS was calculated for all included individuals. For the descriptive analysis, the PRS was standardised to the mean and standard deviation (SD) in healthy population controls. The mean standardised PRS (sPRS) in population controls is therefore 0 with an SD of 1. Standardisation facilitates the comparison between different groups. For further analysis in the family-based cohort, the PRS was standardised to the mean and SD in the family-based cohort including both cases and unaffected relatives.

Total BOADICEA score and polygenic load ($\text{BOADICEA}_{\text{FH}}$)

The pedigrees were collected and drawn for all families, including all known first-degree and second-degree relatives of the genotyped individuals. For 25 of the 561 family members affected with breast cancer, the age of breast cancer diagnosis was not known. For these affected family members, the age at diagnosis was assumed to equal the average age of developing breast cancer in the Netherlands (61 years), or the age at last follow up if this was earlier.

Two different scores were calculated for all individuals in the family cohort by the online risk prediction tool BOADICEA[10], the total BOADICEA score and the polygenic load. The total BOADICEA score (hereafter termed BOADICEA$_{LTR}$) is a measure for lifetime breast cancer risk, and incorporates *BRCA1* and *BRCA2* status, age, birth cohort and a polygenic load. The polygenic load in the BOADICEA model is an estimated polygenetic component representing a large number of loci of small effect to capture the residual familial aggregation of breast cancer and is therefore a measure of the breast cancer family history[15]. Calculation of the polygenic load is described previously by Muranen et al.[15]. To avoid confusion between the variables polygenic load and the PRS, the polygenic load is hereafter termed BOADICEA$_{FH}$. The BOADICEA$_{LTR}$ and BOADICEA$_{FH}$ were calculated by simulating an individual to be at an age of 1 year and unaffected (for cases), that is, lifetime risk at birth, given the family history.

Statistical analysis

To define the degree of correlation between the sPRS and the BOADICEA$_{FH}$, the Pearson correlation coefficient was calculated. A Cox-type random effect regression model was used to estimate the association between the sPRS and breast cancer, adjusting by family history, using the BOADICEA$_{FH}$ (FH) as covariate:

$$\lambda(t_{ij}) = u_i \lambda_0(t_{ij}) \exp\left(\beta_1 sPRS_{ij} + \beta_2 FH_{ij}\right) \qquad (1)$$

where t_{ij} is the age at first diagnosis of breast cancer or the age at censoring for member *j* in family *i*. Censoring was done at age of last contact with the family or death. Censoring at the age of diagnosis for other tumours, if present, did not affect the result. $\lambda_0(t_{ij})$ refers to the baseline hazard, which is left completely unspecified (Cox-type model), β_1 is the main effect of interest, the regression coefficient of the sPRS and β_2 is the effect of the BOADICEA$_{FH}$. In comparing affected to unaffected relatives, it is important to adjust for different numbers of affected versus unaffected relatives per family. We therefore added a family specific random effect *u>0* in our model, shared by the members of the same family. This unobserved heterogeneity shared within families was assumed to follow a gamma distribution.

To evaluate the potential of the sPRS on the reclassification of breast cancer risk, we constructed a new individual breast cancer risk score based on both the BOADICEA$_{LTR}$ and the estimated effect of the sPRS with the model defined by expression 1. Namely, since BOADICEA$_{LTR}$ is defined as the probability of experiencing breast cancer before age 80 years, the new score is calculated as the distribution function at 80 of a Cox proportional hazard model using BOADICEA$_{LTR}$ as baseline (average risk in the sample) and the sPRS as covariate:

$$BOADICEA_{sPRS} = 1\text{-}(1\text{-}BOADICEA_{LTR})\hat{}(e(\beta_1 * sPRS)) \qquad (2)$$

The sPRS is expected to individualise cancer risk estimates, but not to alter the overall average risk level computed by BOADICEA in the joint sample, that is, the higher risks given to some individuals are expected to be compensated by lower risks in others. For this reason we centred the sPRS at the mean of the whole family cohort.

The risk calculation based on BOADICEA alone ($BOADICEA_{LTR}$) and the new individual breast cancer risk score ($BOADICEA_{sPRS}$) were compared for all individuals in the family-based cohort to define the change in risk category and thus advice for breast cancer surveillance according to three different guidelines, NICE[17], NCCN[16] and IKNL[18] (Table S2).

Statistical significance was established at 5%, analysis was performed using R version 3.4.1[25].

Results

The analysis of the ORIGO cohort included 357 breast cancer cases and 327 population controls. The analysis of the family-based cohort included 323 breast cancer cases and 262 unaffected relatives from 101 families. Unaffected relatives derived from 49 of these 101 families.

Descriptive analysis

Virtually all breast cancers were invasive in both cohorts, and second breast cancers were more prevalent in familial cases (Table 1). In both the ORIGO and family-based cohort, the sPRS was on average higher in cases than in controls (Table 2). The unaffected relatives in the family-based cohort had on average a higher sPRS in comparison with ORIGO cases and controls. The mean sPRS for sporadic cases was 0.35 (SD=0.92), and in the family-based cohort, the mean sPRS was 0.70 (SD=0.90) and 0.53 (SD=0.95) for the affected and unaffected relatives respectively. In the family-based cohort, the sPRS was higher for cases with two invasive breast tumours in comparison with cases with one breast tumour (invasive/in situ), with a mean sPRS of 0.66 (SD=0.89) and 0.89 (SD=0.93) respectively. The distributions of the sPRS in both cohorts are shown in Figure 1. Information about the 95% Confidence Interval (CI) and Standard Error (SE) in different groups are shown in Table 2.

Table 1: Characteristics of all included individuals

		ORIGO cohort		Family-based cohort	
		cases	controls	cases	unaffected relatives
Number		357	327	323	262
Age	**Mean (SD)**	56 (10)	46 (14)	51 (11)	62 (17)
	Range	23-84	18-90	26-90	17-94
Country of origin	**The Netherlands**	357	327	317	249
	Hungary	-	-	6	14
First breast tumour	**Invasive (%)**	313 (88)	-	317 (98)	-
	DCIS (%)	32 (9)	-	4 (1)	-
	Unknown (%)	12 (3)	-	2 (1)	-
Second breast tumour	**Invasive (%)**	19 (5)	-	51 (16)	-
	DCIS (%)	2 (1)	-	4 (1)	-
	Unknown (%)	0 (0)	-	5 (2)	-
Family score	**BOADICEA$_{FH}$ (SD)**	-	-	1.03 (0.40)	1.05 (0.39)
	BOADICEA$_{LTR}$ (SD)	-	-	0.23 (0.07)	0.23 (0.06)

Abbreviations: BOADICEA$_{FH}$, Breast cancer family history score; BOADICEA$_{LTR}$, Breast cancer lifetime risk at age 80; DCIS, ductal carcinoma in situ.

Table 2: Descriptive analysis 161-SNP PRS

Group	Mean sPRS	SD sPRS	SE sPRS	n	95% CI	
					lower limit	upper limit
Family breast cancer cases	0.70	0.90	0.05	323	0.60	0.80
1 breast tumour	0.66	0.89	0.05	267	0.55	0.76
2 breast tumours	0.89	0.93	0.12	56	0.65	1.13
Unaffected relatives	0.53	0.95	0.06	262	0.41	0.64
ORIGO cases	0.35	0.92	0.05	357	0.26	0.45
Population controls	0.00	1.00	0.06	327	-0.11	0.11

Abbreviations: PRS, polygenic risk score; SNP, single nucleotide polymorphism; sPRS, standardised PRS.

Correlation

Further analyses were performed only for the family-based cohort. A weak but statistically significant positive correlation was detected between the BOADICEA$_{FH}$ (measure of the family history) and the sPRS. The Pearson correlation coefficient was 0.103, 95% confidence interval (CI)=0.022-0.183, P=0.013, which means that 1.1% of the variance in the sPRS is explained by the BOADICEA$_{FH}$. Larger correlation was found in the unaffected relatives (correlation coefficient 0.153, 95%CI=0.032-0.269, P=0.013). No evidence of correlation was found in family cases only (correlation coefficient 0.057, 95%CI=-0.052-0.165, P=0.306).

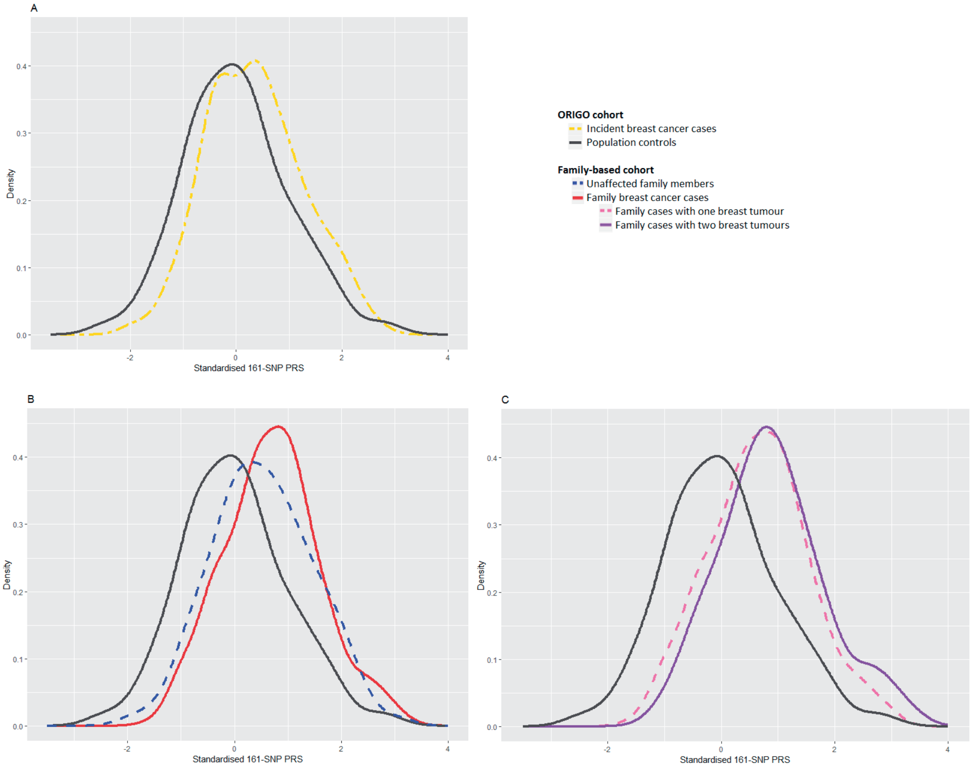

Figure 1: Distribution of the standardised 161-SNP PRS

The standardised 161-SNP PRS was plotted against the density in the different cohorts. (A) incident breast cancer cases and population controls from the ORIGO cohort; (B) population controls from the ORIGO cohort, breast cancer cases and unaffected relatives from the family-based cohort; (C) population controls from the ORIGO cohort, breast cancer cases with one and two primary breast tumours from the family-based cohort.

Abbreviations: PRS, polygenic risk score; SNP, single nucleotide polymorphism

Cox-type random effects modelling

The sPRS should not be directly combined with the BOADICEA$_{LTR}$ because the PRS is a part of the familial relative risk, captured by BOADICEA by its polygenic component, the BOADICEA$_{FH}$. For this reason, adjustment was made by the BOADICEA$_{FH}$ in the association analysis, using model (1). Furthermore, adjusting for the BOADICEA$_{FH}$ helps to correct for ascertainment bias. The BOADICEA$_{FH}$ was calculated for cases assuming they were at age 1 year and unaffected. Consequently controls have, in our sample, a larger BOADICEA$_{FH}$ than cases. Hence, adding the BOADICEA$_{FH}$ as a covariate in the model indirectly corrects the oversampling of cases of our design. Within the family-based cohort, the sPRS was significantly associated with breast cancer, conferring a hazard ratio (HR) of 1.16 (95%CI=1.03-1.28; P=0.026) per SD. No statistical significant association was found without adjustment, HR 1.10, 95%CI= 0.98-1.23, P=0.122.

PRS-based individualised risk score

To calculate a PRS-based breast cancer risk score (BOADICEA$_{sPRS}$), the individual sPRS was combined with the BOADICEA$_{LTR}$. Both risk scores for each individual in the family-based cohort are plotted against each other in Figure 2. This resulted in a change in breast cancer lifetime risk for all individuals. We evaluated the proportions of individuals that would fall in another risk management category, given risk cut-off levels from three different clinical guidelines. Risk management changed for 19.8%, 14.7%, and 11.5% of women under the IKNL[18], NICE[17], and NCCN[16] guidelines, respectively (Table 3). The percentage of family cases and unaffected relatives who changed to a lower or higher risk category based on these guidelines are shown in Table S3. Examples of the change in breast cancer risk category are shown for individuals in three pedigrees in Figure 3 and Table S4._

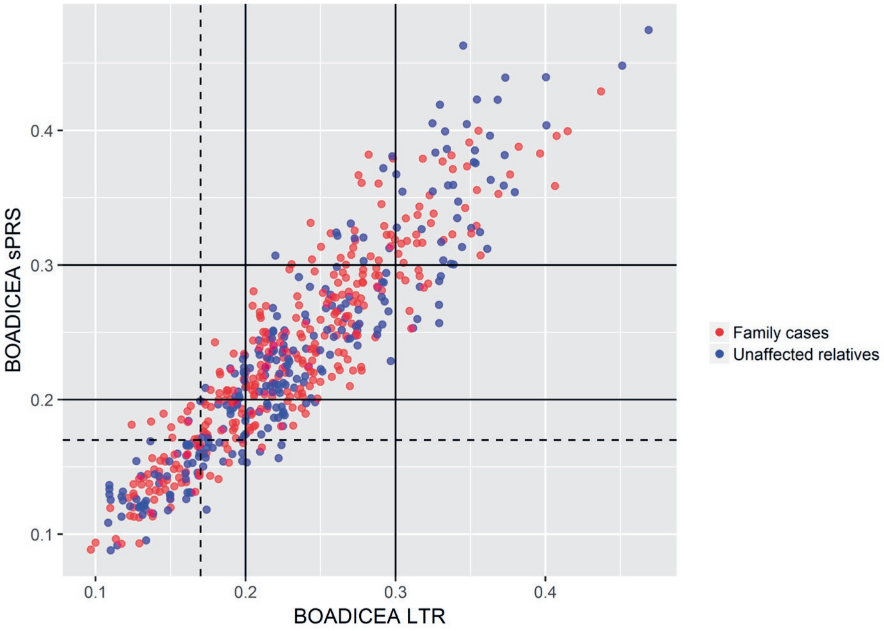

Figure 2: Change in breast cancer lifetime risk score

For every individual, BOADICEA$_{sPRS}$ was plotted against BOADICEA$_{LTR}$. The dotted lines represent the 17% breast cancer lifetime risk cut-off level. The solid lines represent the 20% and 30% breast cancer lifetime risk cut-off levels.

Abbreviations: BOADICEA$_{sPRS}$, 161-SNP PRS based breast cancer lifetime risk score; BOADICEA$_{LTR}$, breast cancer lifetime risk at age 80, based on BOADICEA alone.

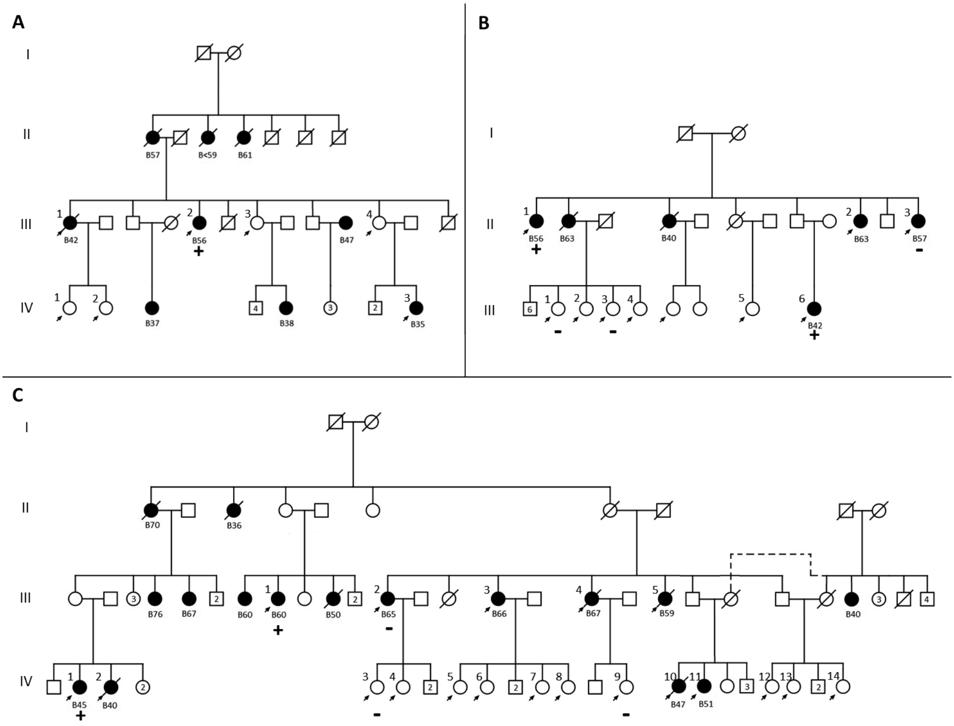

Figure 3: Risk management change for 11 women from three pedigrees
Risk changes are based on the Dutch IKNL screening guideline[18] (Table S2). An arrow indicates that a woman has been genotyped. Generations in the pedigree are numbered with I, II, III and IV.
Based on the individual BOADICEA$_{sPRS}$ score, 11 individuals will change to a higher (+) or lower (-) risk category compared to the BOADICEA$_{LTR}$ score and will receive other breast screening surveillance.
Abbreviations: B, breast cancer; BOADICEA$_{sPRS}$, 161-SNP PRS based breast cancer lifetime risk score; BOADICEA$_{LTR}$, breast cancer lifetime risk at age 80, based on BOADICEA alone; IKNL, Netherlands Comprehensive Cancer Organisation.

Table 3: Breast cancer risk category change in the family-based cohort

Lifetime risk		IKNL[18]		NICE[17]		NCCN[16]	
BOADICEA$_{LTR}$	BOADICEA$_{sPRS}$	N	% change	N	% change	N	% change
<17%	<17%			108	10.7%		
<17%	>17%			13			
17-30%	17-30%			317	15.5%		
17-30%	<17%			24			
17-30%	>30%			34			
<20%	<20%	175	14.2%			175	14.2%
<20%	>20%	29				29	
>20%	>20%					343	10.0%
>20%	<20%					38	
20-30%	20-30%	220	24.7%				
20-30%	<20%	38					
20-30%	>30%	34					
>30%	>30%	74	16.9%	74	16.9%		
>30%	<30%	15		15			
overall change			19.8%		14.7%		11.5%

Following the Dutch IKNL guideline, cut off levels of 20% and 30% represent low, moderate and high risk categories. Following the NICE guideline, 17% and 30% represent low, moderate and high risk categories. Following the NCCN guideline, 20% represent a cut off level for the high risk category.
Abbreviations: BOADICEA$_{LTR}$, breast cancer lifetime risk at age 80, based on BOADICEA alone; BOADICEA$_{sPRS}$, 161-SNP PRS based breast cancer lifetime risk score; IKNL, Netherlands Comprehensive Cancer Organisation; NCCN, National Comprehensive Cancer Network; NICE, National Institute for Health and Care Excellence

Discussion

Polygenic risk scores, derived from a combination of disease-associated SNPs, are gaining importance as predictive factor for a range of disease phenotypes, including breast cancer[26]. All discovered breast cancer SNPs to date explain 18% of the familial relative risk[4]. Here, we use a PRS based on these SNPs, to show the potential clinical utility within high-risk breast cancer families. While most studies use population controls as a reference group[2, 8, 12, 13], we used the healthy relatives of breast cancer cases as a reference to make it more compatible with clinical practice in Family Cancer Clinics. Similar to population-based case-control studies[2, 12, 13], we found that the PRS was significantly associated with breast cancer within high-risk breast cancer families. In addition, the PRS may change breast screening recommendations in a substantial proportion of women from these families, according to currently used screening guidelines[16-18]. For incompleteness of data on ER-status, we did not calculate PRSs predictive for ER-positive or ER-negative disease[5, 27]. While breast cancer screening guidelines are mainly based on overall breast cancer risk, some guidelines suggest discussing the use of chemoprevention with women at high risk of breast cancer[16, 17]. We expect these ER-specific PRSs, similar to the overall PRS, to individualise these discussions within these families.

Some studies have described an association between the PRS and contralateral breast cancer[8, 28]. In agreement with this, we found the average sPRS in women diagnosed with two primary breast cancers in our family cohort to be higher in comparison with women with one breast cancer (similarly in ORIGO cases, Figure S1 and Table S5). Thus, the PRS may be helpful managing contralateral breast cancer risk and guide the choice for treatment or risk reducing mastectomy.

The family-based cohort used in our study was not part of the cohort used to discover the breast cancer associated SNPs by GWAS, while the ORIGO cohort was[3, 4]. A notable finding in our family-based cohort was that unaffected relatives of familial breast cancer cases had on average a higher sPRS than ORIGO incident breast cancer cases, not selected by family history. This may be due to our selection of families with multiple cases of breast cancer, since SNPs of this PRS are expected to cluster in breast cancer families. Moreover, the mean sPRS we calculated for ORIGO cases was lower than found in a large population-based study[2]. Since we found no evidence for substructures in the ORIGO cohort (Figure S1 and Table S5), this effect is probably due to the relatively small number of ORIGO cases included in this study.

Three previous studies have also genotyped breast cancer cases and their unaffected relatives[7, 15, 29]. These studies found an association with breast cancer as well, but effect-sizes are difficult to compare because of differences in methodology and cohort selection criteria. Furthermore, these studies used a much smaller number of SNPs to calculate the PRS. Li et al[7] analysed a prospective dataset, and concluded that their 24-SNP PRS could have altered clinical management in up to 23% of women, regarding an MRI screening-threshold of 20% breast cancer lifetime risk. Evans et al.[29] performed a case-control study of women attending a familial risk clinic, and showed that their 18-SNP PRS moved 52% of the controls without a pathogenic variant in *BRCA1* or *BRCA2* to a different lifetime risk category based on the NICE guideline[17, 29].

In our study, we adopted a conditional approach for association analysis because of the large heterogeneity between the families. Although our use of the $BOADICEA_{FH}$ adjusts for family history, the HR is probably still underestimated given the strong selection criteria used in our study. Of note, this $BOADICEA_{FH}$ is not a true family score in a clinical sense, given the retrospective nature of our family cohort. In clinical practice the risk scores are only calculated for unaffected family members, while in this study, we derive the $BOADICEA_{FH}$ also for cases, assuming they were at age 1 and unaffected. With this definition, controls have, in general, a larger BOADICEA_{FH} than cases. Hence, adding the BOADICEA_{FH} as a covariate in the model indirectly corrects the oversampling of cases of our design. The same definition of the BOADICEA_{FH} is also used when computing BOADICEA_{LTR} and the new individual score BOADICEA_{sPRS}, given by expression (2).

We found that 1.1% of the variance in the sPRS is explained by the BOADICEA_{FH}. Given that 18% of the familial relative risk for breast cancer is explained by the currently known SNPs, this is lower than expected. Nonetheless, other studies have also found a weak correlation or no correlation at all between the PRS and the BOADICEA_{FH} or total BOADICEA score[12, 15]. Thus BOADICEA appears to be a poor predictor of the PRS, underscoring the value of measuring the PRS for every individual in the family instead of using an estimated PRS based on the total family history.

It is estimated that a large number of SNPs just below the level of genome-wide significance, combined with the currently used 161 SNPs, are able to explain about 41% of the familial relative risk[4]. Addition of these SNPs could potentially further refine risk prediction and improve the discriminatory power of the PRS. Studies are now ongoing to find the best performing PRS, including also these SNPs. Khera et al.[30] found that a PRS of 5218 SNPs associated with breast cancer at a significance level of $<5.10^{-4}$, combined with age, had the best performance based on the area under the receiver-operator curve. Mavaddat et al.[31] used a hard-thresholding approach to include 313 SNPs at a significance level of $<10^{-5}$. A further improvement for breast cancer risk prediction could come from information on pathogenic variants in non-BRCA high- or moderate-risk breast cancer genes (e.g. *PALB2, CHEK2, ATM*). Pathogenic variants in these genes are found in approximately 4-6% of women affected with breast cancer[32, 33]. Recently, the BOADICEA model has been extended with incorporation of the effects of truncating variants in *CHEK2, PALB2* and *ATM* and the 313-SNP based PRS to calculate breast cancer lifetime risks[34]. A limitation of our study is that we had no ethical approval to test *CHEK2*, *PALB2* and *ATM* in the studied families. Extrapolating from expected prevalences of pathogenic variants in these genes, we estimate the total percentage of individuals that would have changed to another risk category by addition of the PRS to be 3-4% higher than the 20% we report here.

In summary, we showed that the PRS based on the most recently discovered breast cancer SNPs can be used for breast cancer risk prediction within high-risk breast cancer families. Individualising breast cancer risk preciction by adding the individual 161-SNP PRS to family history-based risk prediction may change screening recommendation in up to 20% of the individuals in these families. While this study illustrates the importance of clinical applicability of the PRS, our results must be interpreted with caution. The HR obtained in this family cohort cannot be translated directly to the clinic as the effect-size must be validated in another larger familial breast cancer cohort. Further evaluation, preferably in prospective settings, will be needed.

Acknowledgements and funding

This work was supported by the Dutch Cancer Society (KWF), grants UL2009-4388 and UL2014-7473.

We would like to thank Prof. D.F. Easton (University of Cambridge, United Kingdom) for critical review of the manuscript. We would like to thank M.E. Braspenning (Leiden University Medical Center, the Netherlands) for drawing all pedigrees from the family-based cohort. We would like to thank dr. S. Böhringer and R.L.M. Tissier (Leiden University Medical Center, the Netherlands) for helpful discussions on statistics at the beginning of the project.

Disclosure of potential conflict of interests

The authors declare no potential conflicts of interest

References

1. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet (London, England)*. Oct 27 2001;358(9291):1389-99. doi:10.1016/s0140-6736(01)06524-2
2. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *JNatlCancer Inst*. 5/2015 2015;107(5)Not in File. doi:djv036 [pii];10.1093/jnci/djv036 [doi]
3. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *NatGenet*. 4/2013 2013;45(4):353-2. Not in File. doi:ng.2563 [pii];10.1038/ng.2563 [doi]
4. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. Oct 23 2017;doi:10.1038/nature24284
5. Lilyquist J, Ruddy KJ, Vachon CM, Couch FJ. Common Genetic Variation and Breast Cancer Risk - Past, present, and future. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. Jan 30 2018;doi:10.1158/1055-9965.epi-17-1144
6. Kuchenbaecker KB, McGuffog L, Barrowdale D, et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *Journal of the National Cancer Institute*. Jul 01 2017;109(7)doi:10.1093/jnci/djw302
7. Li H, Feng B, Miron A, et al. Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. *Genetics in medicine : official journal of the American College of Medical Genetics*. May 12 2016;doi:10.1038/gim.2016.43
8. Sawyer S, Mitchell G, McKinley J, et al. A role for common genomic variants in the assessment of familial breast cancer. *JClinOncol*. 12/10/2012 2012;30(35):4330-4336. Not in File. doi:JCO.2012.41.7469 [pii];10.1200/JCO.2012.41.7469 [doi]
9. Cintolo-Gonzalez JA, Braun D, Blackford AL, et al. Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. *Breast cancer research and treatment*. Jul 2017;164(2):263-284. doi:10.1007/s10549-017-4247-z
10. Lee AJ, Cunningham AP, Kuchenbaecker KB, Mavaddat N, Easton DF, Antoniou AC. BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *British journal of cancer*. Jan 21 2014;110(2):535-45. doi:10.1038/bjc.2013.730
11. Cuzick J, Brentnall AR, Segal C, et al. Impact of a Panel of 88 Single Nucleotide Polymorphisms on the Risk of Breast Cancer in High-Risk Women: Results From Two Randomized Tamoxifen Prevention Trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Mar 2017;35(7):743-750. doi:10.1200/jco.2016.69.8944
12. Dite GS, MacInnis RJ, Bickerstaffe A, et al. Breast Cancer Risk Prediction Using Clinical Models and 77 Independent Risk-Associated SNPs for Women Aged Under 50 Years: Australian Breast Cancer Family Registry. *Cancer EpidemiolBiomarkers Prev*. 12/16/2015 2015;Not in File. doi:1055-9965.EPI-15-0838 [pii];10.1158/1055-9965.EPI-15-0838 [doi]

13. Shieh Y, Hu D, Ma L, et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast cancer research and treatment*. Oct 2016;159(3):513-25. doi:10.1007/s10549-016-3953-2
14. van Veen EM, Brentnall AR, Byers H, et al. Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. *JAMA oncology*. Apr 1 2018;4(4):476-482. doi:10.1001/jamaoncol.2017.4881
15. Muranen TA, Mavaddat N, Khan S, et al. Polygenic risk score is associated with increased disease risk in 52 Finnish breast cancer families. *Breast cancer research and treatment*. Aug 2016;158(3):463-9. doi:10.1007/s10549-016-3897-6
16. NCCN. Clinical Practice Guidelines in Oncology; Breast Cancer Screening and Diagnosis. 2017. Available from: https://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf. Accessed April, 2018
17. NICE. National Institute for Health and Care Excellence: Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer. 2013. Available from: www.nice.org.uk/guidance/cg164. Accessed April, 2018;
18. IKNL. Netherlands Comprehensive Cancer Organisation: Oncoline Mammacarcinoom. 2017. Available from: www.oncoline.nl/richtlijn/item/index.php?pagina=/richtlijn/item/pagina.php&richtlijn_id=885. Accessed April, 2018;
19. de Bock GH, Schutte M, Krol-Warmerdam EM, et al. Tumour characteristics and prognosis of breast cancer patients carrying the germline CHEK2*1100delC variant. *Journal of medical genetics*. Oct 2004;41(10):731-5. doi:10.1136/jmg.2004.019737
20. Oldenburg RA, Kroeze-Jansema KH, Houwing-Duistermaat JJ, et al. Genome-wide linkage scan in Dutch hereditary non-BRCA1/2 breast cancer families identifies 9q21-22 as a putative breast cancer susceptibility locus. *Genes, chromosomes & cancer*. Nov 2008;47(11):947-56. doi:10.1002/gcc.20597
21. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nature methods*. Dec 4 2011;9(2):179-81. doi:10.1038/nmeth.1785
22. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*. Jun 2009;5(6):e1000529. doi:10.1371/journal.pgen.1000529
23. Deelen P, Menelaou A, van Leeuwen EM, et al. Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'. *European journal of human genetics : EJHG*. Nov 2014;22(11):1321-6. doi:10.1038/ejhg.2014.19
24. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature*. Oct 28 2010;467(7319):1061-73. doi:10.1038/nature09534
25. R_Core_Team_(2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
26. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature reviews Genetics*. Jul 2016;17(7):392-406. doi:10.1038/nrg.2016.27

27. Milne RL, Kuchenbaecker KB, Michailidou K, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nature genetics*. Dec 2017;49(12):1767-1778. doi:10.1038/ng.3785
28. Robson ME, Reiner AS, Brooks JD, et al. Association of Common Genetic Variants With Contralateral Breast Cancer Risk in the WECARE Study. *Journal of the National Cancer Institute*. Oct 1 2017;109(10)doi:10.1093/jnci/djx051
29. Evans DG, Brentnall A, Byers H, et al. The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case-control study. *Journal of medical genetics*. Feb 2017;54(2):111-113. doi:10.1136/jmedgenet-2016-104125
30. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*. Sep 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z
31. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American journal of human genetics*. Dec 5 2018;doi:10.1016/j.ajhg.2018.11.002
32. Buys SS, Sandbach JF, Gammon A, et al. A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer*. Jan 13 2017;doi:10.1002/cncr.30498
33. Couch FJ, Shimelis H, Hu C, et al. Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA oncology*. Sep 01 2017;3(9):1190-1196. doi:10.1001/jamaoncol.2017.0424
34. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in medicine : official journal of the American College of Medical Genetics*. Jan 15 2019;21(8):1708-1718. doi:10.1038/s41436-018-0406-9

Supplementary methods

Quality Control

For the ORIGO cohort quality control was performed as part of association studies conducted by the Breast Cancer Association Consortium (BCAC)[1, 2]. To summarise the thresholds used, individuals were excluded when they were genotypically not female, overall call-rate was <95%, low or high heterozygosity ($P<1x10^{-6}$), first-degree relatives determined by identity-by-state estimates or in the case of ancestry outliers by multidimensional scaling. SNPs were excluded with call rates <95% or deviation from Hardy-Weinberg equilibrium in controls at $P<1x10^{-7}$.

For the family-based cohort, quality control was performed with Plink version 1.7[3, 4], which excluded 14342 SNPs with a call rate below 98%. For the remaining SNPs, there was no deviation from Hardy-Weinberg equilibrium in controls at $P<1x10^{-3}$. In total 27 individuals were excluded of which 19 individuals with a call rate below 96% and 6 individuals because of another degree of relatedness than expected based on identity-by-state estimates and pedigree information. Two individuals were genotypically not female and were excluded from further analysis.

Multidimensional scaling was performed to determine clustering of families, including the Hungarian families. There were no different clusters for families, therefore we could also include the Hungarian families.

Supplementary figures and tables

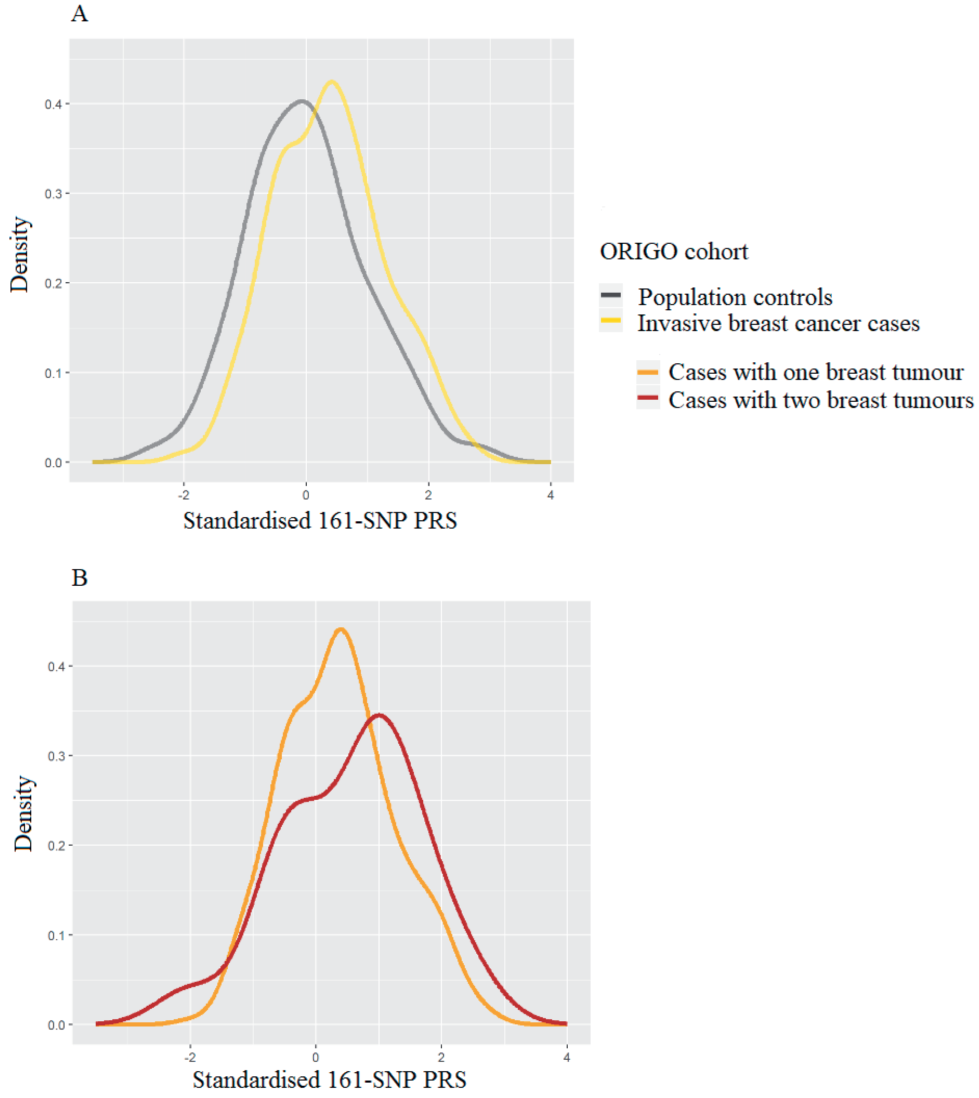

Figure S1: Distribution of the standardised 161-SNP PRS
The standardised 161-SNP PRS was plotted against the density in subgroups of the ORIGO cohort; (A) invasive breast cancer cases and population controls, (B) cases with one versus two breast tumours.

Table S1: 161 breast cancer associated SNPs used for calculating the Polygenic Risk Score[2]
See online material

Table S2: Dutch breast screening guideline (IKNL)[5]

	Low (<2)	Moderate (RR: 2-3)	High (RR: >3)
Life Time Risk	<20%	20-30%	>30%
Start screening	50 yr	40 yr	35 yr
Physical examination	-	-	+
Mammography	population screening*	<50 yr annual >50 yr population screening*	<60 yr annual >60 yr population screening*
MRI	-	-	-

*Biannual mammography

Table S3: Change in risk category for family breast cancer cases and unaffected relatives

	IKNL[5]		NICE[6]		NCCN[7]	
	Lower	**Higher**	**Lower**	**Higher**	**Lower**	**Higher**
Family breast cancer cases	7.4%	12.1%	5.0%	10.8%	5.0%	4.6%
Unaffected relatives	11.1%	9.2%	8.8%	4.6%	8.4%	5.3%

Percentages are based on the total number of family breast cancer cases and unaffected relatives, 323 and 262 respectively.
Following the Dutch IKNL guideline, cut off levels of 20% and 30% represent low, moderate and high risk categories. Following the NICE guideline, 17% and 30% represent low, moderate and high risk categories. Following the NCCN guideline, 20% represent a cut off level for the high risk category.

Table S4: Risk scores from individuals shown in Figure 3

	Individual	Standardised 161-SNP PRS	BOADICEA$_{LTR}$	BOADICEA$_{sPRS}$
Family A	**III-1**	0.62	0.35	0.37
	III-2	0.98	0.29	0.33
	III-3	0.86	0.40	0.44
	III-4	0.07	0.40	0.40
	IV-1	-1.54	0.24	0.20
	IV-2	-0.77	0.24	0.22
	IV-3	0.09	0.20	0.21
Family B	**II-1**	2.14	0.23	0.30
	II-2	0.20	0.32	0.33
	II-3	-0.66	0.32	0.29
	III-1	-0.62	0.20	0.18
	III-2	0.56	0.20	0.21
	III-3	-1.06	0.20	0.17
	III-4	0.81	0.20	0.22
	III-5	-0.56	0.23	0.21
	III-6	2.70	0.22	0.31
	III-7	0.16	0.21	0.22
Family C	**III-1**	1.06	0.27	0.31
	III-2	-0.50	0.31	0.29
	III-3	0.34	0.27	0.28
	III-4	0.12	0.27	0.28
	III-5	0.86	0.26	0.29
	IV-1	1.21	0.26	0.31
	IV-2	0.52	0.25	0.27
	IV-3	-1.42	0.21	0.18
	IV-4	-0.36	0.22	0.21
	IV-5	1.08	0.22	0.25
	IV-6	-0.27	0.22	0.21
	IV-7	0.95	0.22	0.25
	IV-8	0.63	0.22	0.24
	IV-9	-1.41	0.23	0.19
	IV-10	0.17	0.30	0.31
	IV-11	0.28	0.31	0.32
	IV-12	-0.57	0.29	0.27
	IV-13	-0.14	0.29	0.29
	IV-14	-0.42	0.29	0.28

161-SNP PRS, Polygenic Risk Score based on 161 breast cancer associated SNPs; BOADICEA$_{LTR}$, breast cancer lifetime risk at age 80, based on BOADICEA alone; BOADICEA$_{sPRS}$, 161-SNP PRS based individual breast cancer risk score.

Table S5: Mean and SD for ORIGO incident breast cancer cases subgroups

ORIGO cases subgroup	Number	Standardised 161-SNP PRS	
		Mean	SD
Non-invasive tumour*	44	0.21	0.95
Invasive tumour	313	0.37	0.91
1 invasive breast tumour	294	0.36	0.91
2 invasive breast tumours	19	0.56	1.11

*or unknown invasiveness

161-SNP PRS, Polygenic Risk Score based on 161 breast cancer associated SNPs; SD, Standard Deviation

Supplementary references

1. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *NatGenet*. 4/2013 2013;45(4):353-2. Not in File. doi:ng.2563 [pii];10.1038/ng.2563 [doi]
2. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. Oct 23 2017;doi:10.1038/nature24284
3. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. Sep 2007;81(3):559-75. doi:10.1086/519795
4. Purcell S. PLINK version 1.7. http://pngu.mgh.harvard.edu/purcell/plink/;
5. IKNL. Netherlands Comprehensive Cancer Organisation: Oncoline Mammacarcinoom. 2017. Available from: www.oncoline.nl/richtlijn/item/index.php?pagina=/richtlijn/item/pagina.php&richtlijn_id=885. Accessed April, 2018;
6. NICE. National Institute for Health and Care Excellence: Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer. 2013. Available from: www.nice.org.uk/guidance/cg164. Accessed April, 2018;
7. NCCN. Clinical Practice Guidelines in Oncology; Breast Cancer Screening and Diagnosis. 2017. Available from: https://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf. Accessed April, 2018