The path to individualised breast cancer screening

Lakeman, I.M.M.

The path to individualised breast cancer screening

Inge M.M. Lakeman

The path to individualised breast cancer screening
Thesis, department of Human Genetics and Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands.
ISBN: 978-94-6423-798-6

Cover photo: Jan Lakeman
Cover design: Inge Lakeman
Layout and printing: Vera van Beek – ProefschriftMaken || www.proefschriftmaken.nl

The research presented in this thesis was financially supported by the Dutch Cancer Society (grant number: KWF UL2014-7473).

Financial support for publication of this thesis was kindly provided by the department of Human Genetics of the Leiden University Medical Center, and ChipSoft.

The path to individualised breast cancer screening

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op dinsdag 14 juni 2022

klokke 11.15 uur

door

Inge Margaretha Maria Lakeman

geboren te Hoorn

in 1987

Promotores
prof. dr. P. Devilee
prof. dr. C.J. van Asperen

Co-promotor
dr. M. Rodriguez - Girondo

Promotiecommissie

prof. dr. ir. M.K. Schmidt	
prof. dr. A.G. Uitterlinden	*Erasmus Medical Center*
prof. dr. R. Schmutzler	*University of Cologne*
dr. M.A. Adank	*The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital*

Contents

CHAPTER 1

General introduction

General introduction[2]

Breast cancer burden

Breast cancer is worldwide the most common cancer among women, especially in Western Europe[3], and is responsible for almost 25% of the total cancer burden for women[4]. In 2019 and 2020, respectively 17,148 and 14,935 women were diagnosed with breast cancer in the Netherlands[5]. Most breast cancers are detected by mammographic screening. The remainder by palpation of a breast mass, axillary mass or skin abnormalities[6]. Dependent on the abnormality, an additional ultrasound or biopsy is recommended to differentiate between a benign abnormality (e.g. fibroadenoma, ductal hyperplasia), in situ cancer or invasive breast cancer[7, 8]. In situ cancers are classified as low, medium or high grade by histological features[9]. Classification of invasive breast cancer, which can guide treatment options and estimate prognosis, is based on histological type (pathologic growth pattern), grade and tumour stage. More than 20 histological types of breast cancer are known of which the most common are infiltrating duct carcinomas, no special type (70-80%) and invasive lobular carcinomas (~10%)[10]. Tumour grade is a good prognostic factor and includes microscopic assessment of histologic differentiation (tubule formation, nuclear pleomorphism, and proliferation). Tumour stage combines data on tumour size, nodal status and distant metastasis. The most common sites of distant metastasis include the lung, bone and liver. Important for considering hormone therapy is determination of hormone receptor status of the tumour. The majority of breast tumours, about ~75%, express Estrogen Receptor (ER) and/or Progesterone Receptor (PR). Usually, these hormone receptor-positive tumours are low grade and less aggressive. A minority of roughly 15% of breast tumours have overexpression of human epidermal growth factor 2 receptors (HER2), which predict a favourable response to anti-HER2 therapy. However, these tumours are known to be aggressive and have a poor prognosis. Triple-negative breast cancers (i.e., negative for ER, PR and HER2 amplification) comprise about 10% of all breast tumours, are mostly high grade and have a poor prognosis[10].

Breast cancer screening

The high prevalence of breast cancer in the Netherlands equates to an average lifetime risk of 12-13%[7] and provided a strong rationale for a population-screening program that started in 1990. This program invites women every two years for mammography, starting at age 50 and ending at age 75. At age 50 the average 10-year risk to develop invasive breast cancer is approximately 3%, exceeding the threshold at which screening becomes cost-effective[11]. About 63% of all breast cancers in 2019 were detected in women between 50 and 75 years of age[5]. The program has a compliance rate of around 80% and has been demonstrated to cause a decline in mortality rate of approximately 1.7% each year[12]. However, this mortality benefit has been offset by an increasing breast cancer incidence of about twofold[5]. Whether the reduction of mortality can be fully ascribed

to mammographic screening or to improvements of therapeutic options is still under debate. It could be the combination of early diagnosis and therapy[13]. Mammographic screening led to a decrease in the rate of large tumours, and an increase in the detection of small tumours which may represent overdiagnosis[14]. Overdiagnosis is the detection of tumours that, if left untreated, would not have become clinically relevant, mostly Ductal Carcinoma In Situ (DCIS), a non-invasive form of breast cancer. Currently, 13% of the total breast cancer burden in the Netherlands is due to DCIS, while this was about 3% before the start of population screening[5, 9, 15]. Although the majority of DCIS lesions remain indolent, all DCIS are treated with surgery (mostly breast-conserving)[9, 15]. Besides that surgery is resulting in overtreatment of at least some of these lesions, women are labelled as cancer patients and experience substantial psychological distress, which shows the disadvantages of screening. Furthermore, mammographic screening results in a high number of false-positive results[16, 17]. Women attending biannual mammographic screening at age 50, have a cumulative 10-year risk of about 6% for a false-positive result leading to a biopsy[18].

To summarise, secondary prevention by early detection through mammographic screening can reduce mortality, but at the cost of overdiagnosis and the burden of false-positive results[16-18]. Primary prevention by risk reducing mastectomy is in the Netherlands restricted to women at high risk, mainly for *BRCA1/2* pathogenic variant carriers. Stratification of women according to the risk of developing breast cancer could provide a persuasive rationale for surgical intervention as well as improve efficacy of risk–reduction and screening strategies by tailoring starting age and frequency[19, 20].

Breast cancer risk

BOX 1: definition of breast cancer risk
Clinically, definitions such as low, moderate and high breast cancer risk are often used. However, this can reflect relative or absolute risks. For a given relative risk (RR), absolute risk can vary between countries depending on cancer incidences. Another term often used is lifetime risk, which is the absolute risk of breast cancer over the period of a woman's life. Here, we define moderate risk as RR = 2 to 4, high risk as RR > 4, and low or population risk as RR < 2.

To accurately assess a woman's risk, it is important to take all risk factors into account. Having a positive family history is one of the main risk factors for breast cancer. For women with at least one first-degree relative with breast cancer, the risk for developing breast cancer is on average about two-fold compared with women without such a family history[21]. Approximately 25% of this so-called familial relative risk (FRR) is currently explained by (likely) pathogenic variants in a small number of genes, and a further 18% by the currently

known common low risk variants, mostly single nucleotide polymorphisms (SNPs)[22-25]. It is estimated that another 18% is explained by common low risk variants present on SNP arrays used for genotyping in genome-wide association studies, but these have not yet been individually discovered[26] (Figure 1). Besides the familial relative risk, other risk factors such as mammographic density and lifestyle factors are important as well[27, 28].

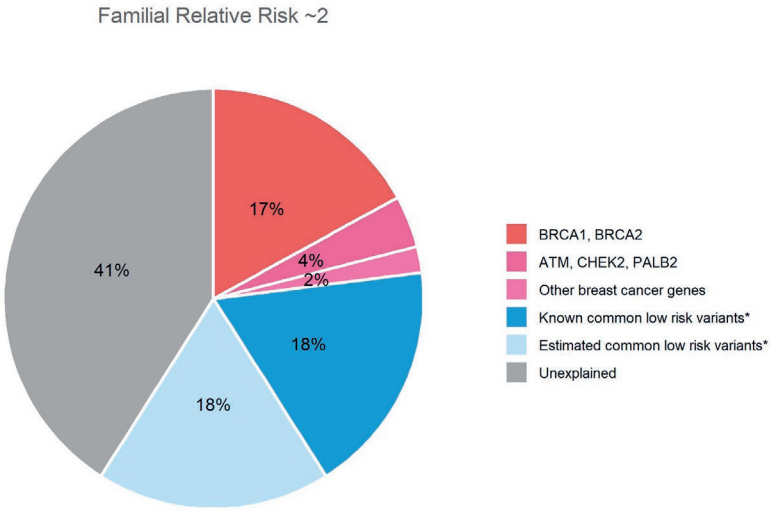


Figure 1. Explained familial relative risk
*For women of European ancestry[26]

Rare genetic variation associated with breast cancer

The definition of "rare" variation is somewhat arbitrary, but is generally taken as to occur in <0.5% of the general population. Indeed, we currently know that some allelic variants in breast cancer susceptibility genes are extremely rare (<0.001%), others moderately rare (~0.1%), or even almost "common" (~1%). In addition, the risks conferred by these variants may vary from less than 2-fold to over 10-fold (Figure 2). Classic linkage analysis in multiple-case families discovered some of these genes, but many were discovered by DNA sequencing of candidate genes. The best-known examples of linkage-detected genes are *BRCA1* and *BRCA2*[29, 30]. Pathogenic variants in either gene, each with a joint allele frequency of ~0.1%, will lead to a high risk of breast and ovarian cancer in women[31, 32]. Other genes, particularly *TP53, PTEN, STK11, CDH1* and *NF1,* were discovered because of their association with typical familial cancer syndromes of which breast cancer is one feature[33-37]. Accordingly, their prevalence in the population is extremely rare. These findings also underscore the pleiotropic effects that some DNA variations display by predisposing to cancers of diverse tissue origin. Yet for most breast cancer genes discovered so far, the most conspicuous "other" cancer with which an association has

been firmly established is ovarian cancer. Another "syndromic" gene is *ATM*; pathogenic variants in *ATM* act in a recessive way to cause ataxia telangiectasia, a neurodegenerative disorder, but heterozygous carriers are at moderately increased risk for breast cancer[38]. The discovery that *BRCA1*, *BRCA2*, and *ATM* are involved in DNA damage repair, and that *BRCA2* is a Fanconi anaemia gene (FANCD1)[39], suggested that other DNA repair genes might also confer breast cancer susceptibility. Sequence analysis of these candidates then led to the discovery of *CHEK2, BARD1, PALB2, NBN,* and *RAD51D*[40-44] as breast cancer genes, although evidence is sometimes limited to specific variants in populations of specific ethnic background[43]. Breast cancer risks in these five genes are generally moderate, with the exception of loss-of-function variants in *PALB2,* which can lead to breast cancer risks comparable to *BRCA2*[43, 45].

There is a long list of genes, including *BRIP1, FANCC, FANCM, MEN1, MRE11A, PPM1D, RAD50, RAD51B, RECQL,* and *XRCC2,* for which an association with breast cancer has been reported in a few studies. Until recently, however, replication in sufficiently large samples of cases and controls and establishment of effect-sizes was still lacking. In 2021, two large population-based case-control studies were published[46, 47] which defined the association of genes often present on commercial breast cancer gene panels with breast cancer risk and provided the most precise risk estimates to date. As expected, robust associations were found for truncating variants in the five well known breast cancer genes, *BRCA1, BRCA2, PALB2, CHEK2*, and *ATM*[46, 47]. Furthermore, truncating variants in *BARD1, RAD51C,* and *RAD51D* were also significantly associated with breast cancer risk in both studies[46, 47], although Hu et al.[47] only detected an association with a ER-negative and triple-negative breast cancer for these genes . An association with truncating variants in respectively *TP53* with overall breast cancer[46], and *CDH1* with ER-positive breast cancer[47] was only found in one of the studies[46, 47]. Modest evidence was demonstrated for an association with truncating variants in *NF1*, *PTEN* and *MSH6*, particularly in ER-negative subtypes. Despite the large sample size, for some genes there is still no consensus about the association with breast cancer risk[46]. A long-standing issue is whether the Lynch syndrome genes (*MLH1, MSH2, MSH6, PMS2*) and *MUTYH* are associated with breast cancer risk. Interpretation of breast cancer incidence in studies of Lynch syndrome families is complicated due to various biases (e.g., ascertainment). Of the lynch syndrome genes, *MSH6* seems to have the highest probability of being associated with breast cancer risk[46, 48]. More detailed discussions on the association of gene variants and breast cancer and the corresponding risks can be found in reviews by Wendt et al., Easton et al., and Graffeo et al.[43, 44, 49].

Allele frequency and corresponding odds ratio for truncating pathogenic variants in associated breast cancer genes, adapted from Dorling et al.[46]. Genes shown in red are robustly associated with breast cancer (p-value <0.0001). Genes shown in orange were marginally associated with breast cancer risk (p-value <0.05). The frequency and

corresponding odds ratio for breast cancer associated common low risk variants, included in the PRS_{313}, are shown in blue and are adapted from the study performed by Mavaddat et al.[26]

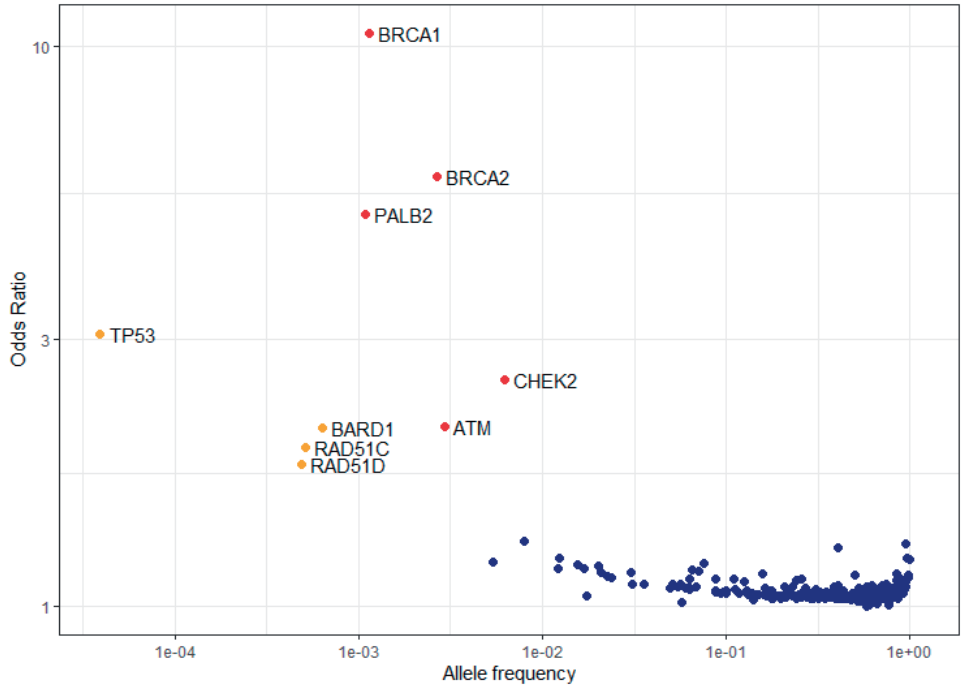


Figure 2. Genetic landscape of breast cancer

Challenges in risk assessment and clinical translation

Once a gene has been robustly associated with breast cancer, other challenges arise that may hamper introduction into the clinic. One is allelic diversity and the notion that different types of variants (e.g., nonsense versus missense changes) might confer different breast cancer risks[43]. For *BRCA1* and *BRCA2*, the effect of mutation-position on the relative risks for breast and ovarian cancer has been firmly established[50]. Furthermore, several missense changes have been identified in *BRCA1* and *BRCA2* that cause much more moderate risks than the typical loss-of-function variants[51, 52]. Conversely, while most pathogenic variants in *ATM* will give an intermediate breast cancer risk, one specific missense mutation (c.7271C>G) seems to reach a higher level of risk. In some studies this risk is even approaching that of *BRCA1/2* pathogenic variants[53, 54]. The presence of allelic diversity in breast cancer genes also highlights the difficulties we are still having with establishing pathogenicity for each variant. This seems straightforward for protein-truncating variants (although exceptions exist[55]), but for many missense and "spliceogenic" variants the impact

on protein function (and, by inference, on cancer risk) is hard to predict. The many *in silico* tools available for this purpose may help classifying these variants, are inexpensive and easy to use, but they still perform modest with respect to clinical standards and, therefore, the predictive power of these tools need to be improved[56]. For some genes, such as *BRCA1* and *BRCA2*, functional assays are developed which show efficacy in variant classification but these are, among other things, time-consuming with a consequence of poor feasibility in daily clinical practice[57]. As a result, many variants detected by sequencing in these genes are still classified as Variants of Uncertain Significance (VUS).

BOX 2: Classification of gene variants

The ACMG has recommended a five-tier classification system, which has been adopted by many countries1. These classes are 1. Benign; 2. Likely Benign; 3. VUS; 4. Likely Pathogenic; 5. Pathogenic. For VUS, the pathogenicity and hence the association with disease risk are unknown, usually because they result in a similarly-shaped amino acid or reside in a part of the gene not essential for its function.

Another challenge is to establish the penetrance of pathogenic variants and the corresponding breast cancer risks with sufficient accuracy. With some exceptions, there is still much uncertainty surrounding the magnitude and precision of the risks conferred by pathogenic variants in the genes. Even in the recently performed large gene-panel studies, the confidence intervals of the associated risks remain wide[46, 47]. One problem underlying this issue is ascertainment bias in the sample used in the analyses. Patient series consisting mostly of women with a positive family history are almost certainly overestimating risk due to enrichment of other risk factors. This is especially true for tumour syndrome genes, investigation of which is usually triggered by the syndrome criteria. For example, the penetrance of *TP53* variants was initially estimated to be very high[58]. But with the introduction of gene panel sequencing, pathogenic variants in *TP53* were also reported in families who do not fulfil the classical criteria of Li-Fraumeni Syndrome[59]. These families show older ages of onset of breast cancer[60], suggesting lower penetrance of at least some *TP53* pathogenic variants. This is consistent with recent estimates of the prevalence of pathogenic germline *TP53* variants in the general population[61], which are also much higher than expected on the basis of the prevalence of Li-Fraumeni Syndrome alone. Furthermore, although with a large confidence interval, Dorling et al. found an OR of approximately 3 for TP53 truncating and missense variants, which is lower than initially demonstrated[58]. The other problem is the rarity of variants, which necessitates the analysis of very large case-control series in order to sufficiently narrow down confidence intervals of risk estimates. For this reason, we have reasonably good breast cancer risk estimates for the 1100delC variant in *CHEK2*, which occurs in ~0.5% of the general population in Europe[62, 63] and the USA[63, 64], but not for most other, much rarer variation in this gene. Even in the recently performed large gene panel studies[46, 47], wide confidence intervals of the

risk are often found. To establish an odds ratio of 2 with a 95% confidence interval of 1.4-2.8, conferred by a variant with an allele frequency of 0.01%, would require genotyping 100,000 cases and 100,000 controls. Larger numbers are needed for lower risks and lower allele frequencies. One way around this problem is to perform burden-type association studies, in which different variants are lumped together on the assumption that their impact on protein function is identical. This is an accepted approach for protein-truncating variants[46, 47], but is problematic for missense changes.

Gene panel studies – non-*BRCA1/2* genes

Gene panel sequencing (GPS) has become a diagnostic reality in cancer genetics. Due to the lower costs and improving data quality, it became possible to test multiple genes in addition to *BRCA1* and *BRCA2* in a single assay, driven by a desire to explain familial clustering of breast cancer in more families and thus impact clinical management. As explained above, the frequency of pathogenic variants found in clinic-based series of familial cases is dependent on the selection criteria of the families included. The highest frequencies, up to 10%, of pathogenic variants are still found in the *BRCA1* and *BRCA2* genes in familial breast cancer cases[65-67] compared to ~2.5% in population-based cases[46, 47]. Pathogenic variants in non-*BRCA1/2* genes are found in 3.7-6.2% of the familial cases[64-69]. The highest frequencies of pathogenic variants in non-*BRCA1/2* genes are found in *CHEK2, ATM* and *PALB2*[64]. However, this increased diagnostic yield comes at the expense of a large proportion of detected VUS, which poses a significant clinical problem. Gene panel studies have found a VUS in 13.6-41.6% of the cases[65, 67, 68, 70]. This means that for every pathogenic variant found in a case, 2 to 3 cases with a VUS are detected. Furthermore, gene panels may contain many genes for which the relevance to breast cancer is unknown or uncertain, as outlined above. Due to these uncertainties, most test-results of commercial gene panels do not translate well into cancer risk assessment. Even the relatively well-defined cancer risks conferred by *BRCA1* and *BRCA2* are influenced by mutation position and mutation class, as well as by other genetic factors, non-genetic exposures, and lifestyle factors[52, 71-73]. Therefore, the gain in clinical utility of testing genes for which evidence of their association with breast cancer is still ill-defined, remains limited[43, 74].

Common low risk variants and Polygenic Risk Scores

Since 2005, Genome-Wide Association Studies (GWAS), using SNP arrays and very large case-control samples, have enabled the identification of common low risk variants for breast cancer[25]. Collaborative groups such as the Breast cancer Association Consortium (BCAC), have identified ~180 common low risk variants associated with breast at genome wide significance level ($1x10^{-8}$)[23]. The first substantial batch of variants was found by the Collaborative Oncologic Gene environment Study (COGS) in 2013, coordinated by BCAC, which was subsequently confirmed and extended by combining with other GWAS data[75]. Another 65 loci were detected after the introduction of the OncoArray, a SNP array with

a much denser SNP coverage than COGS[23]. Some of the associated variants are more strongly associated with ER-negative or ER-positive subtypes of breast cancer[23, 76]. These initially 180 known associated variants explain 18% of the familial relative risk for breast cancer, but a much greater proportion (~40%) can be explained when variants that can be reliably imputed from the OncoArray data were included[23, 26]. Because many of these are expected to be relatively rare (<5%) and/or of very small effect-sizes, very large case-control studies are needed to reach genome-wide significance levels of association. More recent large pooled GWAS discovered already 38 novel breast cancer susceptibility loci at genome wide significance level[77, 78], although some of these loci are only associated with certain breast cancer subtypes.

The breast cancer associated common low risk variant alleles are distributed normally throughout the general population. This means that, in contrast to pathogenic variants in breast cancer susceptibility genes, all individuals in the population carry a certain number of risk alleles, with most individuals carrying the average number. Individually, these risk alleles confer a very small increase in breast cancer risk, but their joint effect may be a substantially higher[22]. In the absence of evidence of clear interactions between variants[22, 79], a simple log-additive (or multiplicative) model combines all variants into a single Polygenic Risk Score (PRS).

Many different PRS for breast cancer have been published in recent years. As published previously[2], Table 1 presents the effect sizes of published PRS until January 2019. Most studies presented here have generated PRS for overall unilateral breast cancer[22, 26, 80-89], few have addressed ER-status-specific PRS-models with the use of subtype-specific odds ratios of certain SNPs[26, 90]. Subtype-specific PRS can potentially be useful to guide clinical management for chemoprevention and other prevention strategies. Some studies[91-93] have used a PRS to predict contralateral breast cancer, and others studied the PRS as risk modifier in rare gene mutation carriers (*BRCA1, BRCA2, CHEK2*)[94-96]. The number of common low risk variants, their allele frequencies and effect-sizes determine the discriminatory and predictive power of a PRS. Predictive power of a PRS is usually expressed as Odds or Hazard Ratio (OR, HR) per Standard Deviation (SD) unit of the distribution; discriminatory power is often assessed by the area under the curve (AUC). The number of variants included in a PRS is not strongly correlated with the overall effect-size or the AUC. This is because the variants detected in the earliest studies, although smaller in number, generally have higher effect-sizes than those detected more recently in larger studies with more statistical power. Including large numbers of variants at lower than genome-wide significance thresholds may increase predictive power of the PRS, but at the expense of being less specific[26].

Table 1: Effect size and AUC of Polygenic Risk Scores

Study[a]		Cases	Controls	Variants	Effect size				AUC
					Measurement	Overall BC	ER-positive BC	ER-negative BC	
	Unilateral breast cancer								
McCarthy et al. 2015[80]	prospective cohort; women referred for breast biopsy	74	390	12	per SD	OR=2.30[1.06-4.99] P=0.035			0.685 [0.620-0.750] (including age)
Dite et al. 2015[81]	population based case-control; women <50yr	750	405	77	per SD, adjusted for age group	OR=1.46 [1.29-1.64], 2*10-16			0.61[0.58-0.65]
Mavaddat et al. 2015[22]	case-control study	33673	33381	77	per SD	OR=1.55 [1.52-1.58]	OR=1.63[1.60-1.67]	OR=1.45 [1.40-1.49]	0.622 [0.619-0.627]
Naslund-koch et al. 2016[82]	Danish population study	1301	19010	72	5th in comparison to 1st quintile	HR=1.82[1.53-2.18]			NA
Li et al. 2016[89]	Prospective cohort; breast cancer families	1496	2869	24	per SD	HR=1.38[1.22-1.56] P=2.9*10^-7			0.59[0.55-0.63]
Shieh et al. 2016[83]	nested case-control study; mammography screening cohort	486	495	83	4th quartile in comparison to 1st quartile	OR= 2.54[1.69-3.82], P<0.001			0.60 [0.57-0.64]
Muranen et al. 2016[84]	case-control study	1689	1269	75	per SD	OR=1.56 [1.45-1.68] P=9.2E-31			NA
Muranen et al. 2016[84]	Breast cancer families	181	1269	75	per SD	OR=1.82 [1.55-2.13] P=1.8E-13			NA
Maas et al. 2016[85]	prospective cohort; nested case-control study	17171	19862	24; 92	10th decile in comparison to 1th decile	OR=2.79 (24 variants)			0.623 (92 variants)
Cuzick et al. 2017[101]	prospective nested case-control study	359	636	88	Interquartile range	OR=1.37 [1.16-1.79] ; p-value <0.001	OR=1.44 [1.16-1.79] P<0.001	OR=0.99 [0.61-1.61] P=0.10	0.55 [0.51-0.60]
Shieh et al. 2017[90]	nested case-control study; mammography screening cohort	110 (ER-positive)	214	83	per SD		OR=1.58 [1.06-2.36] p=0.02		0.68 [0.61-0.75] p=0.07
Evans et al. 2017[86]	Case-control study; women attending a familial risk clinic	364 (112 *BRCA1/2+*)	1605 (691 *BRCA1/2+*)	18	Interquartile range	Non-carriers: OR=1.55 [1.29-1.87]			Non-carriers: 0.59[0.55-0.63]

Study[a]		Cases	Controls	Variants	Effect size				AUC
					Measurement	Overall BC	ER-positive BC	ER-negative BC	
van Veen et al. 2018[102]	Prospective cohort study	466	8897	18	interquartile range	OR=1.56 [1.38-1.77]			NA
Zhang et al. 2018[87]	nested case-control study	4006	7874	67	4th quartile in comparison to 1st quartile	RR=2.5 [2.2-2.8]			NA
Khera et al. 2018[88]	Case-control study; UK biobank participants	6586	157895	5218	5th quintile in comparison with remainder	OR=2.07 [1.97-2.19], P=3.4*10-159			0.69[0.68-0.69], including age
Mavaddat et al. 2019[26]	Prospective case-control studies	11428	18323	313	per SD	OR= 1.61[1.57-1.65]	OR=1.45 [1.37-1.53]	OR=1.35 [1.27-1.43]	0.630 [0.628-0.651]
Mavaddat et al. 2019[26]	Case-control study; UK biobank	3215	186825	306	per SD	HR=1.59 [1.54-1.64]			
	Unilateral breast cancer in gene mutation carriers								
Muranen et al. 2017[95]	*CHEK2* c.1100delC carriers	39,139 (624 carriers)	40,063 (224 carriers)	74	per SD	carriers: OR=1.59[1.21-2.09], P=0.0008; non-carriers: OR=1.58[1.55-1.62], P<1.0E-10			NA
Kuchenbaecker et al. 2017[94]	*BRCA1* mutation carriers	7,797	7,454	88	per SD	HR=1.14 [1.11-1.17], P=1.8*10-18	HR=1.11 [1.08-1.15], P=3.5*10-13	HR=1.27 [1.23-1.31], P=8.2*10-53	Overall BC: 0.541 [0.530-0.551]
Kuchenbaecker et al. 2017[94]	*BRCA2* mutation carriers	4,330	3,881	88	per SD	HR=1.22 [1.17-1.28], P=7.2*10-20	HR=1.22 [1.16-1.27], P=4.0*10-19	HR=1.15 [1.10-1.20], P=6.8*10-10	Overall BC: 0.566 [0.551-0.581]
	Contralateral breast cancer								
Sawyer et al. 2012[92]	Case-control study; Familial breast cancer cohort	126 contralateral BC	711 unilateral BC	22	4th quartile in comparison with 1st quartile	OR=1.96 [1.17-3.70]			NA
Robson et al. 2017[91]	Population based case-control study, <55yr	1,459 contralateral BC	2,126 unilateral BC	67	4th quartile in comparison with 1st quartile	OR = 1.75 [1.41-2.18]			NA

[a]For unilateral breast cancer, studies are added up to January 2019.

Abbreviations: AUC, Area Under the Curve; BC, Breast Cancer; ER, Estrogen Receptor; SD, Standard Deviation.

For all PRS-models predicting breast cancer, the AUC is modest, i.e., 0.6 – 0.7, but should this alone preclude their application as an individual test to predict if a woman will develop breast cancer or not? A comparison with gene panel testing, which is widely used in the clinic for this purpose, is illustrative. A PRS has been shown to be capable of stratifying women into different risk categories in a clinically meaningful way[22, 89, 92, 94], but the most relevant clinical information of the PRS is in the extreme tails of the distribution. Because these tails concern the general population (as opposed to gene mutation carriers only), the associated attributable risks of the PRS are in fact far greater than that achieved by gene panel testing. For example, the best performing PRS at this moment includes 313 common low risk variants (PRS_{313}) with an association at a p-value threshold two orders below genome-wide significance ($P<10^{-5}$). For this PRS, in the general population, 35% of all breast cancers occur in women in the highest quintile and only 9% of all breast cancers in the lowest quintile[26]. Women in the top 1% of the PRS_{313} are at 4-fold elevated risk relative to population average (95% CI 3.34-4.89), a risk-level defined in many countries as 'high'. In comparison, *BRCA1* mutation carriers explain <2% of all breast cancer in Western Caucasian populations[97] and comprise ~0.1% of the general population. Additional studies have shown that the PRS based on 313 variants is associated with both contralateral breast cancer in the population[98] and unilateral breast cancer among *BRCA1/2* gene mutation carriers[73]. Implementation research is ongoing to introduce the PRS into clinical genetic testing e.g. in the Netherlands, Germany, France, UK and USA. An example of how individual PRS-testing could aid risk counselling in the setting of familial breast cancer is shown in Figure 3, which highlights how two individuals that would otherwise have received the same risk assessment (sisters in generation IV) on the basis of their identical family history, are clearly classified into distinct risk classes on the basis of their PRS_{313}.

Another potential application of the PRS is in deciding when and how frequent women should undergo breast cancer screening[20, 99]. In most countries running such screening programs, women are offered screening above a certain age, usually between 45 and 50, when their breast cancer risk exceeds a certain cost-effective level. Women in the lowest quintile of the PRS_{313} in fact never reach that threshold, whereas those in the highest quintile will attain this level of risk before age 40 years[26]. A risk-based entry into population-screening, as opposed to the current age-based entry, could therefore be more cost-effective, although the evidence to support this notion has been derived only from modelling studies so far[20, 100].

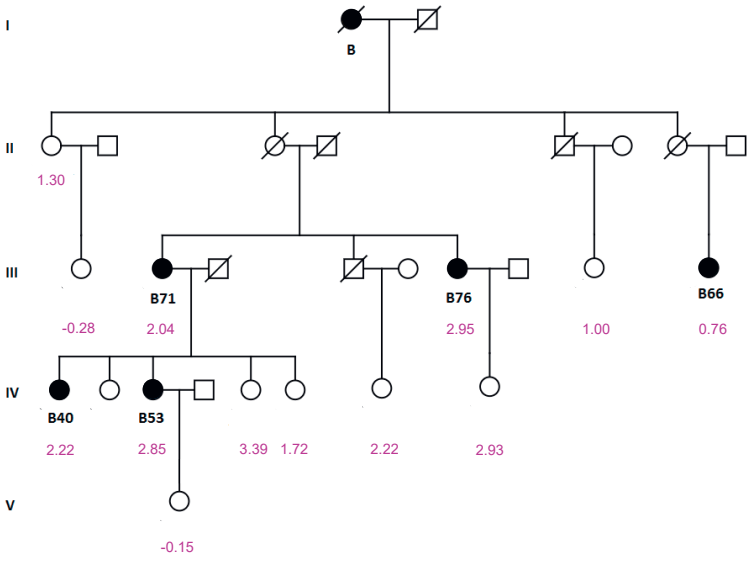


Figure 3. Standardised PRS for breast cancer cases and their female relatives
In this non-*BRCA1/2* breast cancer family, multiple family members were genotyped by SNP array. For all genotyped individuals, the SNP_{313} Polygenic Risk Score (PRS) was calculated. The individual PRS are standardised to population controls in the BCAC dataset (mean=0 and SD=1 in controls). The numbers in the figure are therefore Z-scores of the individual PRS. A higher Z-score indicates a higher breast cancer risk.

A limitation of many PRS is that most variants contained in it are discovered in European-descent populations and their effects cannot be translated directly to other ethnicities. Studies are ongoing to define breast cancer associated variants and evaluate the European-descent derived PRS in non-European populations. Recently, studies performed for the Asian population[103] and Latinas[104], showed similar performance for the PRS as in the European population, but for the African population[105] there was an attenuated effect size. Therefore, caution is needed when using the European-descent derived PRS for women of ancestries for which the effect of the PRS is dissimilar or not yet determined.

Hormonal, environmental and lifestyle risk factors

A number of non-genetic risk factors are presently firmly established as being associated with breast cancer. Besides age, these include physical factors such as body height and weight[106, 107]. For weight, breast cancer risk is dependent on menopausal status. Weight gain and obesity (BMI>30) after menopause are associated with an increase in postmenopausal breast cancer[106]. It is likely that higher oestrogen levels underlie this effect in postmenopausal women[108]. A higher mammographic density due to a high proportion of connective and glandular relative to adipose tissue, leads to a higher risk for breast cancer[27, 28, 109]. Hormonal factors influencing breast cancer risk include the use of oral contraception and hormone replacement therapy (HRT)[110, 111], as well as age at menarche and menopause[112]. Reproductive history (age of first childbirth or nulliparity) may have similar impact on mammary gland biology[28, 113]. The lifestyle factors alcohol use and smoking increase breast cancer risk as well, while physical activity and breastfeeding seems to act protectively[114-116]. Finally, a personal history of benign breast disease also signifies an increased breast cancer risk[28].

Combining risk factors

Since any woman will have only a single certain risk-level at a given moment in time to develop breast cancer over the course of her life, genetic and non-genetic risk factors must somehow combine to define that risk. A major challenge for individual breast cancer risk prediction, therefore, is to design risk calculation models that accommodate all known risk factors, which requires knowledge about the underlying model how they interact. Through the large international consortia such as BCAC, data to design and validate such models are now forthcoming. There are now much more accurate estimates how the PRS can modify the breast cancer risks conferred by pathogenic variants in *BRCA1, BRCA2* and *CHEK2*[73, 94, 95, 117, 118]. This can help inform choices and timing of preventive surgery or chemoprevention. The interaction between the c.1100delC variant in *CHEK2* and the PRS appears to follow a simple multiplicative interaction, but the per SD hazard ratio estimates in *BRCA1* and *BRCA2* pathogenic variant carriers were smaller than those in general population[73]. In *BRCA1* pathogenic variant carriers, the ER-negative PRS showed a much stronger association with breast cancer risk in comparison with the ER-positive

PRS, consistent with the predominant ER-negative tumour subtype in *BRCA1* pathogenic variant carriers[73, 94]. Few studies have been performed on *ATM* and *PALB2* pathogenic variant carriers, but a recent study showed that the effect sizes of the associations were in between those for *BRCA1/2* and *CHEK2*[119]. These issues highlight the complexity of some of these interactions and underscore the necessity of large prospective cohort studies to validate these models. A similar deviation from simple multiplicative interactions has been found for individuals with rare pathogenic variants in more than one breast cancer associated gene[120]. There is limited evidence for interaction between common low risk variants and lifestyle/hormonal factors[121]. Recent studies showed that the effect of these risk factors and the PRS can in general be combined in a multiplicative way[122, 123].

Breast cancer risk prediction models

Currently, predicting whether a healthy woman will develop a primary breast cancer or not is mainly done within clinical genetic services. Women who are worried because of their family history for breast cancer can be referred by their general practitioner to such a clinic; alternatively, breast cancer patients with a clear family history are referred by oncologists or surgeons, also because of the potential impact a gene diagnosis may have for their therapeutic options[124]. At the moment, the major incentive behind these referrals is the possibility to detect a high- or moderate risk variant in one of the breast cancer genes (i.e. *BRCA1*, *BRCA2*, *PALB2, CHEK2, or ATM)*. As set forth above, however, such variants are found in ~10% of all referred families. For women from breast cancer families where no pathogenic variant is found, clinical management is determined based on their lifetime breast cancer risk. The Dutch screening guideline (Table 2) advises women with a risk above 20% based on their family history to perform annual mammography from age 40, and to continue biennial screening at age 50 as part of the national population screening program. An intensified protocol has been designed for women with a risk >30%. Breast cancer risk prediction for healthy relatives is often based on family history alone, although more than 20 risk prediction algorithms known today[125] include other risk factors as well. 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[81, 83, 101, 102]. In one study[89], new breast cancer lifetime risks for women from breast cancer families were calculated by adding the PRS to family-based risk prediction. For up to 23% of the women, screening recommendations, as stipulated by local management guidelines, could alter.

Some well-known risk prediction algorithms are the Gail model, BRCAPRO, Tyrer-Cuzick and the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA). Depending on what the model predicts and for which population, the most appropriate model can be used. The Gail model predicts breast cancer lifetime risks for women older than 35 years and is widely studied and validated. It includes hormonal risk

factors, breast biopsies and affected first degree relatives[125-127]. The Chen model extends this by incorporating mammographic breast density as well[128]. The BRCAPRO model calculates breast cancer lifetime risks and the risk of contralateral breast cancer. The estimation is based on family history (first- and second-degree relatives), the prevalence of *BRCA1* and *BRCA2* pathogenic variants, population incidence rates and pathological markers for breast cancers[127, 129]. The Tyrer-Cuzick model incorporates hereditary (first- and second-degree relatives with breast or ovarian cancer), hormonal and environmental risk factors (age, BMI, menarche, reproductive factors, menopause, and HRT) and pathological variables (breast biopsies and benign breast pathology)[125, 127]. Mammographic density and PRS were recently incorporated in the model[127]. BOADICEA estimates breast cancer lifetime risks and contralateral cancer risks for women with a family history of breast cancer[130]. The model includes tumour pathology characteristics, recent cancer incidences and pathogenic variants in *ATM, BRCA1, BRCA2, CHEK2,* and *PALB2*[131]. For BOADICEA, the family history is not restricted to a number of relatives or a particular degree. The current version, model V5, has been extended to accommodate a broad range of genetic and non-genetic risk factors for breast cancer, adding mammographic density, reproductive factors, age at menarche and menopause, use of hormones, BMI, body height, alcohol use and 4 different PRS including the PRS_{313} to the previous version[132]. In the new version, V6, available in February 2022, breast and ovarian cancer population incidences of the Netherlands will be added. Unsurprisingly, the potential for risk stratification was the greatest when all risk factors were used for risk prediction. Of all factors, the PRS had the largest contribution in risk stratification. Without knowledge of the genetic status of a woman for the rare genes, or family history, the lifetime breast cancer risk varied from 2.8% for the lowest, to 30.6% for the highest percentile of the PRS[132]. The model assumes that the risk factors and the PRS_{313} act multiplicatively, consistent with evidence from previous studies[123]. Similarly, the assumption that the PRS_{313} combines multiplicatively with the effects of rare truncating variants in the five breast cancer genes will need validation. Finally, the current BOADICEA model uses population breast cancer risks of several countries but UK risk factor distributions and therefore may require tailoring for application in other populations. The BOADICEA model is incorporated in the user-friendly web interface CanRisk[133] and externally validated[134]. Within clinical genetic services of the Netherlands, CanRisk is already used by some clinicians for risk prediction in families where no pathogenic variant is found, but currently mostly only family history is included as variable.

Outline of this thesis

The main objective of work presented in this thesis was to explore the clinical utility of the Polygenic Risk Score (PRS) based on breast cancer associated common low risk variants for individual breast cancer risk prediction. It did so by generating knowledge about the PRS in the Dutch general population and in clinic-based breast cancer families, as well

as in a large international population of *BRCA1/2* pathogenic variant carriers. The results will support implementation of comprehensive risk prediction by using CanRisk in the clinic, and may help women to make more informed choices about their optimal clinical management.

Table 2: Dutch screening guideline

	Low (RR: <2)	Moderate (RR: 2-3)	High (RR: >3)
Lifetime risk	<20%	20-30%	>30%
Start screening	50 yr	40 yr	35 yr
Mammography	Population screening	<50 yr annual >50 yr population screening	<60 yr annual >60 yr population screening
MRI	-	-	-

Chapter 2 explores the clinical applicability of a 161-variant-based 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 association with breast cancer and the clinical impact of the PRS on risk prediction was investigated for affected and healthy women from these families by determining the potential change in clinical management.

Chapter 3 explores the clinical applicability of the 313-variant-based PRS for risk prediction in a cohort of almost 4,000 familial Dutch breast cancer cases who tested negative for pathogenic variants in *BRCA1/2* and of whom the majority were evaluated in research setting for pathogenic variants in *PALB2, CHEK2*, and *ATM*. The clinical impact of addition of the PRS on breast cancer risk prediction by BOADICEA based on family history and pathogenic variant carrier status was investigated by determining the potential change in clinical management. In Appendix 1, this study is used as illustration to discuss the situation with regard to the review by the Medical Ethical Committees of multicentre research in the Netherlands that is not covered by the Dutch medical research involving human subjects act (wet medisch-wetenschappelijk onderzoek met mensen, WMO) [article in Dutch].

Chapter 4 assesses the clinical validity of the 313-variant-based PRS by determining the association between this PRS and breast cancer in the Dutch population. Furthermore, we validated the risk prediction algorithm BOADICEA by exploring the discriminative ability of an individualised 10-year breast cancer risk score based on the PRS and several known risk factors. We also assessed how a risk-based approach of population-based screening could have impacted breast cancer detection rates in our study cohort.

In chapter 5, we investigated whether the 313-variant-based PRS for breast cancer is associated with contralateral breast cancer risk among women with pathogenic variants in *BRCA1 or BRCA2* and explored the implications for contralateral breast cancer risk prediction for these women.

In chapter 6, we summarised the results of our pilot study, the Individualised Breast cancer Risk prediction (IBR) study in which we included unaffected women from breast cancer families where no pathogenic variant is found. The aim of this study was to establish the percentages of women shifting to another risk category with comprehensive risk prediction (CRP) calculated using CanRisk, based on family history, the PRS_{313} and lifestyle/hormonal risk factors compared to the current family history-based risk prediction. Furthermore, the psychosocial impact of this new CRP will be assessed and described by Bredart et al. (*manuscript submitted*).

In chapter 7 we conclude with a general discussion about our main findings and future perspectives for implementation of CRP for breast cancer in the clinic.

References

1. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics*. May 2015;17(5):405-24. doi:10.1038/gim.2015.30
2. Lakeman IMM, Schmidt MK, van Asperen CJ, Devilee P. Breast Cancer Susceptibility—Towards Individualised Risk Prediction. journal article. *Current Genetic Medicine Reports*. June 01 2019;7(2):124-135. doi:10.1007/s40142-019-00168-5
3. The Global Cancer Observatory. World Health Organization International Agency for Research on Cancer - Breast cancer fact sheet. Accessed October 12, 2021. https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf
4. The Global Cancer Observatory. World Health Organization International Agency for Research on Cancer - Cancer fact sheet. Accessed October 12, 2021. https://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf
5. IKNL. Cijfers over kanker, Nederlandse kankerregistratie. Accessed October 12, 2021. http://www.cijfersoverkanker.nl
6. Esserman LJ, Shieh Y, Rutgers EJ, et al. Impact of mammographic screening on the detection of good and poor prognosis breast cancers. *Breast cancer research and treatment*. Dec 2011;130(3):725-34. doi:10.1007/s10549-011-1748-z
7. Kennisinstituut van Federatie Medisch Specialisten. Richtlijn Borstkanker. . Accessed October 13, 2021. https://richtlijnendatabase.nl/richtlijn/borstkanker/algemeen.html
8. Bevers TB, Helvie M, Bonaccio E, et al. Breast Cancer Screening and Diagnosis, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network : JNCCN*. Nov 2018;16(11):1362-1389. doi:10.6004/jnccn.2018.0083
9. van Seijen M, Lips EH, Thompson AM, et al. Ductal carcinoma in situ: to treat or not to treat, that is the question. *British journal of cancer*. Aug 2019;121(4):285-292. doi:10.1038/s41416-019-0478-6
10. Tsang JYS, Tse GM. Molecular Classification of Breast Cancer. *Adv Anat Pathol*. Jan 2020;27(1):27-35. doi:10.1097/pap.0000000000000232
11. van der Waal D, Verbeek AL, den Heeten GJ, Ripping TM, Tjan-Heijnen VC, Broeders MJ. Breast cancer diagnosis and death in the Netherlands: a changing burden. *European journal of public health*. Apr 2015;25(2):320-4. doi:10.1093/eurpub/cku088
12. Otto SJ, Fracheboud J, Looman CW, et al. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. *Lancet (London, England)*. Apr 26 2003;361(9367):1411-7. doi:10.1016/s0140-6736(03)13132-7
13. Tabár L, Dean PB, Chen TH, et al. The incidence of fatal breast cancer measures the increased effectiveness of therapy in women participating in mammography screening. *Cancer*. Feb 15 2019;125(4):515-523. doi:10.1002/cncr.31840

14. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *New England Journal of Medicine*. 2016;375(15):1438-1447. doi:10.1056/NEJMoa1600249
15. Badve SS, Gökmen-Polar Y. Ductal carcinoma in situ of breast: update 2019. *Pathology*. Oct 2019;51(6):563-569. doi:10.1016/j.pathol.2019.07.005
16. Bleyer A. Screening mammography: update and review of publications since our report in the New England Journal of Medicine on the magnitude of the problem in the United States. *Academic radiology*. Aug 2015;22(8):949-60. doi:10.1016/j.acra.2015.03.003
17. Myers ER, Moorman P, Gierisch JM, et al. Benefits and Harms of Breast Cancer Screening: A Systematic Review. *Jama*. Oct 20 2015;314(15):1615-34. doi:10.1001/jama.2015.13183
18. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Annals of internal medicine*. Oct 18 2011;155(8):481-92. doi:10.7326/0003-4819-155-8-201110180-00004
19. Burton H, Chowdhury S, Dent T, Hall A, Pashayan N, Pharoah P. Public health implications from COGS and potential for risk stratification and screening. *Nature genetics*. Apr 2013;45(4):349-51. doi:10.1038/ng.2582
20. Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. *JAMA oncology*. Nov 1 2018;4(11):1504-1510. doi:10.1001/jamaoncol.2018.1901
21. 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
22. 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]
23. 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
24. 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]
25. 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
26. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
27. Nazari SS, Mukherjee P. An overview of mammographic density and its association with breast cancer. *Breast Cancer*. May 2018;25(3):259-267. doi:10.1007/s12282-018-0857-5

28. Nelson HD, Zakher B, Cantor A, et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. *Annals of internal medicine*. May 1 2012;156(9):635-48. doi:10.7326/0003-4819-156-9-201205010-00006
29. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science (New York, NY)*. Oct 7 1994;266(5182):66-71.
30. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature*. Dec 21-28 1995;378(6559):789-92. doi:10.1038/378789a0
31. Antoniou A, Pharoah PD, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *American journal of human genetics*. May 2003;72(5):1117-30. doi:10.1086/375033
32. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Apr 10 2007;25(11):1329-33. doi:10.1200/jco.2006.09.1066
33. Nelen MR, Padberg GW, Peeters EA, et al. Localization of the gene for Cowden disease to chromosome 10q22-23. *Nature genetics*. May 1996;13(1):114-6. doi:10.1038/ng0596-114
34. Hearle N, Schumacher V, Menko FH, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. *Clinical cancer research : an official journal of the American Association for Cancer Research*. May 15 2006;12(10):3209-15. doi:10.1158/1078-0432.Ccr-06-0083
35. Pharoah PD, Guilford P, Caldas C. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology*. Dec 2001;121(6):1348-53.
36. Madanikia SA, Bergner A, Ye X, Blakeley JO. Increased risk of breast cancer in women with NF1. *American journal of medical genetics Part A*. Dec 2012;158a(12):3056-60. doi:10.1002/ajmg.a.35550
37. Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science (New York, NY)*. Nov 30 1990;250(4985):1233-8.
38. Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature genetics*. Aug 2006;38(8):873-5. doi:10.1038/ng1837
39. D' Andrea AD, Grompe M. The Fanconi anaemia/BRCA pathway. *Nature Reviews Cancer*. 1/1/2003 2003;3(1):23-34. In File.
40. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nature genetics*. May 2002;31(1):55-9. doi:10.1038/ng879
41. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature genetics*. Feb 2007;39(2):165-7. doi:10.1038/ng1959
42. Bogdanova N, Feshchenko S, Schurmann P, et al. Nijmegen Breakage Syndrome mutations and risk of breast cancer. *International journal of cancer*. Feb 15 2008;122(4):802-6. doi:10.1002/ijc.23168

43. Easton DF, Pharoah PD, Antoniou AC, et al. Gene-panel sequencing and the prediction of breast-cancer risk. *NEnglJMed*. 6/4/2015 2015;372(23):2243-2257. Not in File. doi:10.1056/NEJMsr1501341 [doi]
44. Wendt C, Margolin S. Identifying breast cancer susceptibility genes - a review of the genetic background in familial breast cancer. *Acta oncologica (Stockholm, Sweden)*. Jan 3 2019:1-12. doi :10.1080/0284186x.2018.1529428
45. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. *New England Journal Of Medicine*. 8/7/2014 2014;371(6):497-506. In File.
46. Dorling L, Carvalho S, Allen J, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *The New England journal of medicine*. Jan 20 2021;doi:10.1056/NEJMoa1913948
47. Hu C, Hart SN, Gnanaolivu R, et al. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *The New England journal of medicine*. Feb 4 2021;384(5):440-451. doi:10.1056/NEJMoa2005936
48. Lu HM, Li S, Black MH, et al. Association of Breast and Ovarian Cancers With Predisposition Genes Identified by Large-Scale Sequencing. *JAMA oncology*. Aug 16 2018;doi:10.1001/jamaoncol.2018.2956
49. Graffeo R, Livraghi L, Pagani O, Goldhirsch A, Partridge AH, Garber JE. Time to incorporate germline multigene panel testing into breast and ovarian cancer patient care. *Breast cancer research and treatment*. Dec 2016;160(3):393-410. doi:10.1007/s10549-016-4003-9
50. Rebbeck TR, Mitra N, Domchek SM, et al. Modification of BRCA1-Associated Breast and Ovarian Cancer Risk by BRCA1-Interacting Genes. *Cancer research*. 9/1/2011 2011;71(17):5792-5805. In File.
51. Shimelis H, Mesman RLS, Von Nicolai C, et al. BRCA2 Hypomorphic Missense Variants Confer Moderate Risks of Breast Cancer. Article. *Cancer research*. Jun 2017;77(11):2789-2799. doi:10.1158/0008-5472.can-16-2568
52. Moghadasi S, Meeks HD, Vreeswijk MP, et al. The BRCA1 c. 5096G>A p.Arg1699Gln (R1699Q) intermediate risk variant: breast and ovarian cancer risk estimation and recommendations for clinical management from the ENIGMA consortium. *Journal of medical genetics*. Jan 2018;55(1):15-20. doi:10.1136/jmedgenet-2017-104560
53. Southey MC, Goldgar DE, Winqvist R, et al. PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. *Journal of medical genetics*. Dec 2016;53(12):800-811. doi:10.1136/jmedgenet-2016-103839
54. Bernstein JL, Teraoka S, Southey MC, et al. Population-based estimates of breast cancer risks associated with ATM gene variants c.7271T>G and c.1066-6T>G (IVS10-6T>G) from the Breast Cancer Family Registry. *Hum Mutat*. 11/2006 2006;27(11):1122-1128. Not in File.
55. Thompson ER, Gorringe KL, Rowley SM, et al. Reevaluation of the BRCA2 truncating allele c.9976A > T (p.Lys3326Ter) in a familial breast cancer context. *Scientific Reports*. 2015 2015;5:14800. In File.
56. Padilla N, Moles-Fernández A, Riera C, et al. BRCA1- and BRCA2-specific in silico tools for variant interpretation in the CAGI 5 ENIGMA challenge. *Human mutation*. Sep 2019;40(9):1593-1611. doi:10.1002/humu.23802

57. Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *J Exp Clin Cancer Res*. Mar 4 2020;39(1):46. doi:10.1186/s13046-020-01554-6
58. Chompret A, Brugieres L, Ronsin M, et al. P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *British journal of cancer*. Jun 2000;82(12):1932-7. doi:10.1054/bjoc.2000.1167
59. O'Shea R, Clarke R, Berkley E, et al. Next generation sequencing is informing phenotype: a TP53 example. *Familial cancer*. Jan 2018;17(1):123-128. doi:10.1007/s10689-017-0002-1
60. Rana HQ, Gelman R, LaDuca H, et al. Differences in TP53 Mutation Carrier Phenotypes Emerge From Panel-Based Testing. *Journal of the National Cancer Institute*. Feb 26 2018;doi:10.1093/jnci/djy001
61. de Andrade KC, Frone MN, Wegman-Ostrosky T, et al. Variable population prevalence estimates of germline TP53 variants: A gnomAD-based analysis. Article. *Human mutation*. Jan 2019;40(1):97-105. doi:10.1002/humu.23673
62. Consortium TCBCC-C. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from ten studies. *Am J Hum Genet*. 6/2004 2004;74(6):1175-1182. In File.
63. Schmidt MK, Hogervorst F, van Hien R, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Aug 10 2016;34(23):2750-60. doi:10.1200/jco.2016.66.5844
64. 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
65. Tung N, Battelli C, Allen B, et al. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. *Cancer*. 1/1/2015 2015;121(1):25-33. Not in File. doi:10.1002/cncr.29010 [doi]
66. Desmond A, Kurian AW, Gabree M, et al. Clinical Actionability of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Risk Assessment. *JAMA Oncol*. 10/2015 2015;1(7):943-951. Not in File. doi:2425836 [pii];10.1001/jamaoncol.2015.2690 [doi]
67. Lerner-Ellis J, Khalouei S, Sopik V, Narod SA. Genetic risk assessment and prevention: the role of genetic testing panels in breast cancer. *ExpertRevAnticancer Ther*. 2015 2015;15(11):1315-1326. Not in File. doi:10.1586/14737140.2015.1090879 [doi]
68. Kapoor NS, Curcio LD, Blakemore CA, et al. Multigene Panel Testing Detects Equal Rates of Pathogenic BRCA1/2 Mutations and has a Higher Diagnostic Yield Compared to Limited BRCA1/2 Analysis Alone in Patients at Risk for Hereditary Breast Cancer. *AnnSurgOncol*. 10/2015 2015;22(10):3282-3288. Not in File. doi:10.1245/s10434-015-4754-2 [doi]
69. Thompson ER, Rowley SM, Li N, et al. Panel Testing for Familial Breast Cancer: Calibrating the Tension Between Research and Clinical Care. *JClinOncol*. 1/19/2016 2016;Not in File. doi:JCO.2015.63.7454 [pii];10.1200/JCO.2015.63.7454 [doi]
70. Susswein LR, Marshall ML, Nusbaum R, et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *GenetMed*. 12/17/2015 2015;Not in File. doi:gim2015166 [pii];10.1038/gim.2015.166 [doi]

71. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *Jama*. Jun 20 2017;317(23):2402-2416. doi:10.1001/jama.2017.7112
72. Rebbeck TR, Friebel TM, Friedman E, et al. Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. *Human mutation*. May 2018;39(5):593-620. doi:10.1002/humu.23406
73. Barnes DR, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in medicine : official journal of the American College of Medical Genetics*. Oct 2020;22(10):1653-1666. doi:10.1038/s41436-020-0862-x
74. Turnbull C, Sud A, Houlston RS. Cancer genetics, precision prevention and a call to action. *Nature genetics*. Sep 2018;50(9):1212-1218. doi:10.1038/s41588-018-0202-0
75. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *NatGenet*. 4/2015 2015;47(4):373-380. Not in File. doi:ng.3242 [pii];10.1038/ng.3242 [doi]
76. 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
77. Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature genetics*. Jun 2020;52(6):572-581. doi:10.1038/s41588-020-0609-2
78. Adedokun B, Du Z, Gao G, et al. Cross-ancestry GWAS meta-analysis identifies six breast cancer loci in African and European ancestry women. *Nat Commun*. Jul 7 2021;12(1):4198. doi:10.1038/s41467-021-24327-x
79. Milne RL, Herranz J, Michailidou K, et al. A large-scale assessment of two-way SNP interactions in breast cancer susceptibility using 46 450 cases and 42 461 controls from the breast cancer association consortium. *Human molecular genetics*. 2014 2014;23(7):1934-1946. In File.
80. McCarthy AM, Keller B, Kontos D, et al. The use of the Gail model, body mass index and SNPs to predict breast cancer among women with abnormal (BI-RADS 4) mammograms. *Breast Cancer Res*. 2015 2015;17:1. Not in File. doi:10.1186/s13058-014-0509-4 [doi];s13058-014-0509-4 [pii]
81. 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]
82. Naslund-Koch C, Nordestgaard BG, Bojesen SE. Common breast cancer risk alleles and risk assessment: A study on 35,441 individuals from the Danish general population. *Annals of oncology : official journal of the European Society for Medical Oncology*. Oct 13 2016;doi:10.1093/annonc/mdw536

83. 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
84. 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
85. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA oncology*. Oct 1 2016;2(10):1295-1302. doi:10.1001/jamaoncol.2016.1025
86. 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
87. Zhang X, Rice M, Tworoger SS, et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. *PLoS medicine*. Sep 2018;15(9):e1002644. doi:10.1371/journal.pmed.1002644
88. 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
89. 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
90. Shieh Y, Hu D, Ma L, et al. Joint relative risks for estrogen receptor-positive breast cancer from a clinical model, polygenic risk score, and sex hormones. *Breast cancer research and treatment*. Nov 2017;166(2):603-612. doi:10.1007/s10549-017-4430-2
91. 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
92. 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]
93. Kramer I, Hooning MJ, Mavaddat N, et al. Breast Cancer Polygenic Risk Score and Contralateral Breast Cancer Risk. *American journal of human genetics*. Nov 5 2020;107(5):837-848. doi:10.1016/j.ajhg.2020.09.001
94. 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
95. Muranen TA, Greco D, Blomqvist C, et al. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. Article. *Genetics in Medicine*. May 2017;19(5):599-603. doi:10.1038/gim.2016.147

96. Barnes D, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in Medicine*. 2020;accepted for publication
97. Ford D, Easton D, Peto J. Estimates of the gene frequency of BRCA1 and its contribution to breast and ovarian cancer incidence. *American journal of human genetics*. 1995 1995;57:1457-1462. In File.
98. Kramer I, Hooning MJ, Mavaddat N, et al. Breast cancer polygenic risk score and contralateral breast cancer risk *American Journal of Human genetics* 2020;(under review)
99. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *The New England journal of medicine*. Jun 26 2008;358(26):2796-803. doi:10.1056/NEJMsa0708739
100. van den Broek JJ, Schechter CB, van Ravesteyn NT, et al. Personalizing Breast Cancer Screening Based on Polygenic Risk and Family History. *Journal of the National Cancer Institute*. Apr 6 2021;113(4):434-442. doi:10.1093/jnci/djaa127
101. 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
102. 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
103. Ho WK, Tan MM, Mavaddat N, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun*. Jul 31 2020;11(1):3833. doi:10.1038/s41467-020-17680-w
104. Shieh Y, Fejerman L, Lott PC, et al. A polygenic risk score for breast cancer in U.S. Latinas and Latin-American women. *Journal of the National Cancer Institute*. Sep 25 2019;doi:10.1093/jnci/djz174
105. Du Z, Gao G, Adedokun B, et al. Evaluating Polygenic Risk Scores for Breast Cancer in Women of African Ancestry. *Journal of the National Cancer Institute*. Mar 26 2021;doi:10.1093/jnci/djab050
106. Lahmann PH, Hoffmann K, Allen N, et al. Body size and breast cancer risk: findings from the European Prospective Investigation into Cancer And Nutrition (EPIC). *International journal of cancer*. Sep 20 2004;111(5):762-71. doi:10.1002/ijc.20315
107. Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *The Lancet Oncology*. Aug 2011;12(8):785-94. doi:10.1016/s1470-2045(11)70154-1
108. Key TJ, Appleby PN, Reeves GK, et al. Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. *Journal of the National Cancer Institute*. Aug 20 2003;95(16):1218-26.

109. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. Jun 2006;15(6):1159-69. doi:10.1158/1055-9965.Epi-06-0034
110. Beral V, Reeves G, Bull D, Green J. Breast cancer risk in relation to the interval between menopause and starting hormone therapy. *Journal of the National Cancer Institute*. Feb 16 2011;103(4):296-305. doi:10.1093/jnci/djq527
111. Hunter DJ, Colditz GA, Hankinson SE, et al. Oral contraceptive use and breast cancer: a prospective study of young women. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. Oct 2010;19(10):2496-502. doi:10.1158/1055-9965.Epi-10-0747
112. Collaborative_Group_on_Hormonal_Factors_in_Breast_Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *The Lancet Oncology*. Nov 2012;13(11):1141-51. doi:10.1016/s1470-2045(12)70425-4
113. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. *Lancet (London, England)*. Jul 20 2002;360(9328):187-95. doi:10.1016/s0140-6736(02)09454-0
114. Hamajima N, Hirose K, Tajima K, et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *British journal of cancer*. Nov 18 2002;87(11):1234-45. doi:10.1038/sj.bjc.6600596
115. Gram IT, Park SY, Kolonel LN, et al. Smoking and Risk of Breast Cancer in a Racially/Ethnically Diverse Population of Mainly Women Who Do Not Drink Alcohol: The MEC Study. *American journal of epidemiology*. Dec 1 2015;182(11):917-25. doi:10.1093/aje/kwv092
116. Pizot C, Boniol M, Mullie P, et al. Physical activity, hormone replacement therapy and breast cancer risk: A meta-analysis of prospective studies. *European journal of cancer (Oxford, England : 1990)*. Jan 2016;52:138-54. doi:10.1016/j.ejca.2015.10.063
117. Mars N, Widén E, Kerminen S, et al. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun*. Dec 14 2020;11(1):6383. doi:10.1038/s41467-020-19966-5
118. Gao C, Polley EC, Hart SN, et al. Risk of Breast Cancer Among Carriers of Pathogenic Variants in Breast Cancer Predisposition Genes Varies by Polygenic Risk Score. *Journal of Clinical Oncology*. 0(0):JCO.20.01992. doi:10.1200/jco.20.01992
119. Gallagher S, Hughes E, Wagner S, et al. Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. *JAMA Network Open*. 2020;3(7):e208501-e208501. doi:10.1001/jamanetworkopen.2020.8501
120. Turnbull C, Seal S, Renwick A, et al. Gene-gene interactions in breast cancer susceptibility. *Human molecular genetics*. Feb 15 2012;21(4):958-62. doi:10.1093/hmg/ddr525

121. Rudolph A, Chang-Claude J, Schmidt MK. Gene-environment interaction and risk of breast cancer. *British journal of cancer*. Jan 19 2016;114(2):125-33. doi:10.1038/bjc.2015.439
122. Rudolph A, Song M, Brook MN, et al. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. *International journal of epidemiology*. Jan 5 2018;doi:10.1093/ije/dyx242
123. Kapoor PM, Mavaddat N, Choudhury PP, et al. Combined Associations of a Polygenic Risk Score and Classical Risk Factors With Breast Cancer Risk. *Journal of the National Cancer Institute*. Mar 1 2021;113(3):329-337. doi:10.1093/jnci/djaa056
124. Pilié PG, Gay CM, Byers LA, O'Connor MJ, Yap TA. PARP Inhibitors: Extending Benefit Beyond BRCA-Mutant Cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research*. Jul 1 2019;25(13):3759-3771. doi:10.1158/1078-0432.Ccr-18-0968
125. 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
126. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*. Dec 20 1989;81(24):1879-86.
127. Kim G, Bahl M. Assessing Risk of Breast Cancer: A Review of Risk Prediction Models. *J Breast Imaging*. Mar-Apr 2021;3(2):144-155. doi:10.1093/jbi/wbab001
128. Chen J, Pee D, Ayyagari R, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute*. Sep 6 2006;98(17):1215-26. doi:10.1093/jnci/djj332
129. Mazzola E, Blackford A, Parmigiani G, Biswas S. Recent Enhancements to the Genetic Risk Prediction Model BRCAPRO. *Cancer informatics*. 2015;14(Suppl 2):147-57. doi:10.4137/cin.S17292
130. 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
131. Lee AJ, Cunningham AP, Tischkowitz M, et al. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. *Genetics in medicine : official journal of the American College of Medical Genetics*. Dec 2016;18(12):1190-1198. doi:10.1038/gim.2016.31
132. 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
133. Carver T, Hartley S, Lee A, et al. CanRisk Tool—A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer Epidemiology Biomarkers & Prevention*. 2021;30(3):469-473. doi:10.1158/1055-9965.Epi-20-1319

134. Pal Choudhury P, Brook MN, Hurson AN, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. *Breast Cancer Research*. 2021/02/15 2021;23(1):22. doi:10.1186/s13058-021-01399-7

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 ($BOADICEA_{FH}$)

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:

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

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 $\text{and } \beta_2$ 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:

$$BOADICEA_{sPRS} = 1-(1-BOADICEA_{LTR})^{(e(\beta_1 * sPRS))} \quad (2)$$

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).

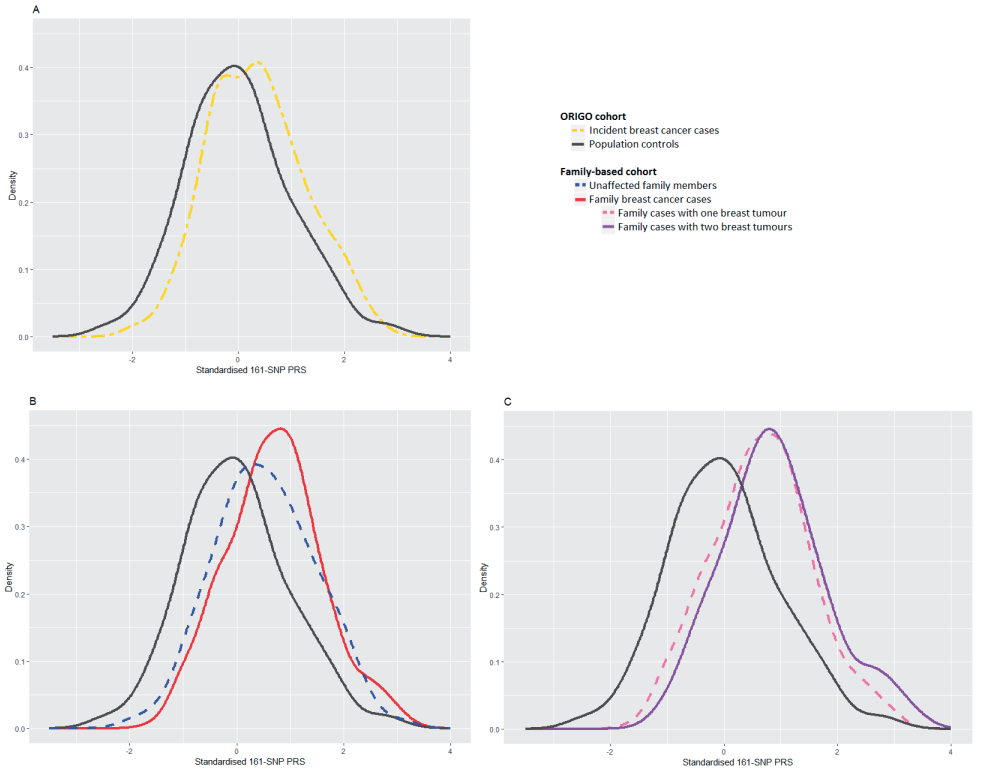


Figure 1: Distribution of the standardised 161-SNP PRS
The standardised 161-SNP PRS was plotted against the density in the different cohorts. (A) incident breast cancer cases and population controls from the ORIGO cohort; (B) population controls from the ORIGO cohort, breast cancer cases and unaffected 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.

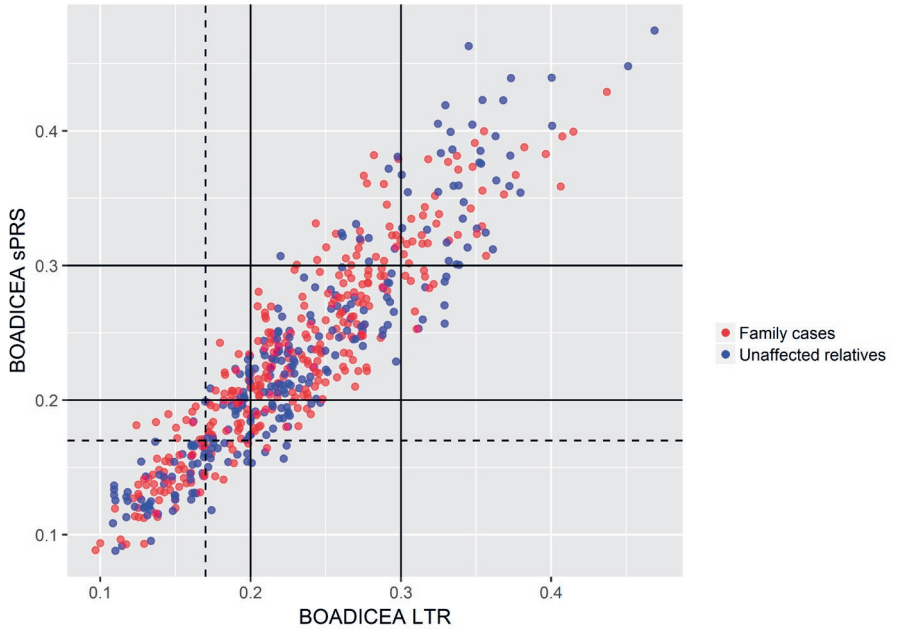


Figure 2: Change in breast cancer lifetime risk score

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.

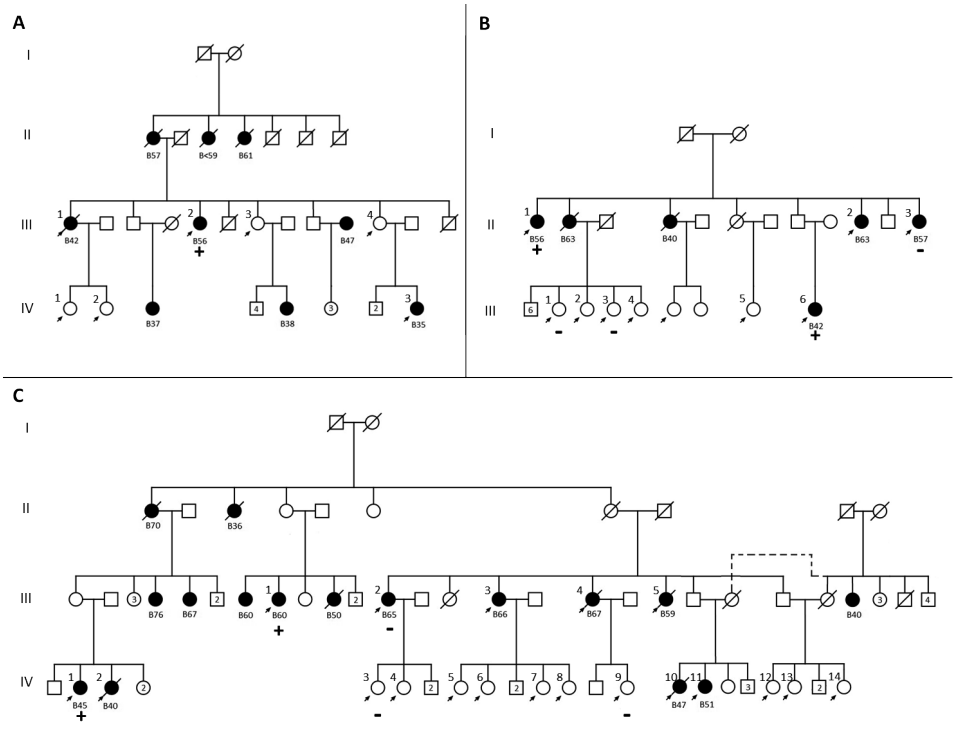


Figure 3: Risk management change for 11 women from three pedigrees
Risk changes are based on the Dutch IKNL screening guideline[18] (Table S2). An arrow indicates that a woman has been genotyped. Generations in the pedigree are numbered with I, II, III and IV.
Based on the individual $BOADICEA_{sPRS}$ 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

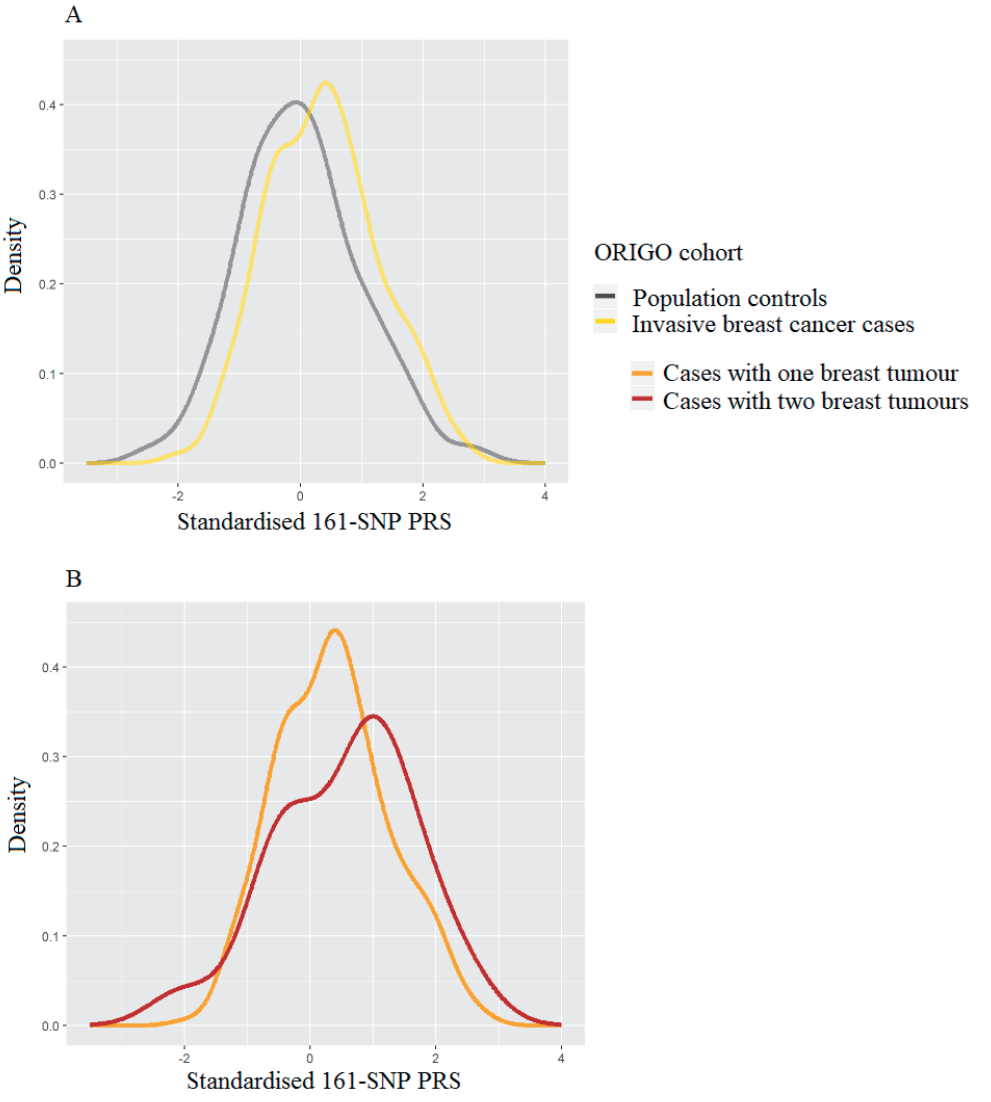


Figure S1: Distribution of the standardised 161-SNP PRS
The standardised 161-SNP PRS was plotted against the density in subgroups of the ORIGO cohort; (A) invasive breast cancer cases and population controls, (B) cases with one versus two breast tumours.

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

CHAPTER 3

3

Clinical applicability of the Polygenic Risk Score for breast cancer risk prediction in familial cases

Inge M.M. Lakeman, Mar Rodriguez-Girondo, Andrew Lee, Nandi Celosse, Merel Braspenning, K. van Engelen, I. van de Beek, A.H. van der Hout, E. Gomez Garcia, A. Mensenkamp, M.G.E.M. Ausems, Maartje Hooning, Muriel A. Adank, Antoinette Hollestelle, Marjanka K. Schmidt, Christi J. van Asperen, and Peter Devilee

Lakeman IMM, et al. Submitted

Abstract

Background: Common low-risk variants are presently not used to guide clinical management of familial breast cancer (BC). We explored the additive impact of a 313-variant-based Polygenic Risk Score (PRS_{313}) relative to standard gene-testing in non-*BRCA1/2* Dutch BC families.

Methods: We included 3,918 BC cases from 3,492 Dutch non-*BRCA1/2* BC families and 3,474 Dutch population controls. The association of the standardised PRS_{313} with BC was estimated using a logistic regression model, adjusted for pedigree-based family history. Family history of controls was imputed for this analysis. Standard errors were corrected to account for relatedness of individuals. Using BOADICEA model version 5, lifetime risks were retrospectively calculated with and without individual PRS_{313}. For 2,586 cases and 2,584 controls, carrier status of pathogenic variants (PVs) in *ATM*, *CHEK2,* and *PALB2* was known.

Results: The family history adjusted PRS_{313} was significantly associated with BC (per SD OR=1.97, 95%CI[1.84-2.11]). Including the PRS_{313} in BOADICEA family-based risk prediction would have changed screening recommendations in up to 27%, 36%, and 34% of the cases according to BC screening guidelines from the USA, UK and the Netherlands (NCCN, NICE, and IKNL), respectively. For the population controls, without information on family history, this was up to 39%, 44%, and 58%, respectively. Among carriers of PVs in known moderate BC susceptibility genes, the PRS_{313} had the largest impact for *CHEK2* and *ATM*.

Conclusions: Our results support the application of the PRS_{313} in risk prediction for genetically uninformative BC families and families with a PV in moderate BC risk genes.

Introduction

Breast cancer (BC) is the most common cancer among women[1]. Current screening strategies to reduce the burden of the disease have several disadvantages, including overdiagnosis[2]. By taking into account all relevant risk factors, personalised estimation of BC risk could help to target preventive measures to those who would benefit the most and to reduce screening for women in the lowest risk categories.

One of the main BC risk factors is having a positive family history of the disease[3]. The familial relative risk of ~2 is partly explained by germline pathogenic variants (PVs) in the BC susceptibility genes *BRCA1/2, PALB2, ATM* and *CHEK2*. Furthermore, another important part is explained by common low-risk variants[4, 5], which, if summarised in a Polygenic Risk Score (PRS), are useful for stratifying the population into different risk categories[5, 6]. A similar stratification of BC risk by the PRS is observed in the familial setting[7-10], providing an opportunity to personalising risk and clinical management for women from BC families who are seen at clinical genetic services. Furthermore, the PRS can be useful in refining risk for women carrying a PV in *BRCA1/2, PALB2, CHEK2,* or *ATM*[11-14]. However, using the PRS for risk prediction is not yet implemented in the practice of genetic counselling for familial BC in the Netherlands.

Currently, risk prediction for women from non-*BRCA1/2* BC families is mainly based on family history, which can be calculated by various risk prediction algorithms[15], such as the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA)[16]. Several studies have shown an improved discriminative power between BC cases and controls by combining the PRS with other risk factors in a BC risk prediction tool[17-20]. Previously, we showed that in a selected group of high risk non-*BRCA1/2* BC families, a 161-variant PRS alone would have led 20% of the women to receive different screening recommendations based on the Dutch screening guideline (Netherlands Comprehensive Cancer Organisation guideline (IKNL))[21]. Currently, the most predictive PRS, based on 313 variants (PRS_{313})[5], is incorporated in the validated, comprehensive risk prediction model BOADICEA[16] that was recently made easily accessible for clinicians through the CanRisk webtool[22].

Here, we explore the clinical applicability of the PRS_{313} for risk prediction in a new cohort of 3,918 familial Dutch BC cases who tested negative in a diagnostic setting for PVs in *BRCA1/2* and of whom the majority were evaluated for PVs in *PALB2, CHEK2*, and *ATM* in a research setting. The clinical impact of the PRS_{313} on BC risk prediction based on family history and PV carrier status was investigated by determining the potential change in clinical management, as stipulated by three currently used guidelines (the National

Comprehensive Cancer Network guideline (NCCN)[23], the National Institute for Health and Care Excellence guideline (NICE)[24], and IKNL[21]).

Materials and Methods

We used the STROBE case-control checklist when writing our report[25].

Study cohorts

Dutch familial BC cases, henceforth "cases", were derived from three different cohorts: the Hereditary Breast and Ovarian cancer study in the Netherlands (HEBON)[26], the Amsterdam Breast Cancer Study-Familial (ABCS-F)[27], and the Rotterdam Breast Cancer Study (RBCS)[28] (Supplementary methods). All three studies included participants who visited a clinical genetic centre in the Netherlands for familial BC counselling. Women with BC who met the following criteria were eligible for this study: 1) family without *BRCA1/2* PVs; 2) available DNA sample or genotyping data; 3) European ancestry based on genotyping data; 4) available pedigree. In total, 3,918 cases were included (Figure S1). All cancers were verified by linkage to the Dutch Cancer Registry and the Pathological Anatomical National Automated Archive (HEBON cases) or by clinical confirmation from medical records in the hospital (ABCS-F and RBCS cases).

In total, 3,474 Dutch population controls of age 18 years or older were included. These controls were healthy female blood donors (ABCS, Oorsprong van borstkanker integraal onderzocht (ORIGO)) or healthy women who were included after DNA diagnostic testing for Cystic Fibrosis carrier status (RBCS)[4, 28] for which age of last follow up was known.

Ethics approval statement

Informed consent was obtained from all included cases, and we received approval for this study of the Medical Ethical Committees of all included centres. All controls were anonymised.

Gene panel

As part of the BRIDGES project, 2,586 cases and 2,584 controls were sequenced for a panel of 34 genes as described elsewhere[29]. For all controls and 2,037 cases, we received results of all included genes. Truncating and missense variants were reported as described previously[29]. In summary, pathogenic truncating variants were defined as frameshift insertions/deletions, stop/gain or canonical splice variants as classified by the Ensembl Variant Effect Predictor[30], with the exception of variants in the last exon of each gene. In our study, we included truncating variants in the last exon of *PALB2*, as this exon encodes an important functional domain and variants in this exon were shown to destabilise

the resulting *PALB2* protein[31]. Missense variants were included if their frequency in the gnomAD database or among the BRIDGES project control dataset[29] was below 0.001. For genes with evidence of an association with BC[29], pathogenicity was reported for missense variants based on the ClinVar archive[32]. For the remaining 549 cases, only pseudo-anonymised results of truncating variants in the three additional BC genes, *ATM, CHEK2*, and *PALB2,* were received, excluding truncating variants in the last exon.

Genotyping and imputation

DNA samples of all included individuals were genotyped for common variants with either the iCOGS[33], OncoArray[4] or Global Screening Array (GSA), containing 211,155, 499,170, and 642,824 Single Nucleotide Polymorphisms, respectively. Genotyping and quality control for the samples genotyped with iCOGS and OncoArray were performed as part of association studies conducted by the Breast Cancer Association Consortium (BCAC)[4, 33]. Genotyping and quality control for the samples genotyped with the GSA array are described in the supplementary methods.

The variants that were not directly genotyped were imputed using the Michigan imputation server[34], using the Haplotype Reference Consortium (HRC) 1.1 reference panel[35] including both the reference panels 1000 Genomes phase 3 and Genome of the Netherlands (GoNL)[36, 37]. In total, 72 of the 313 variants could not be imputed with the HRC1.1 reference panel and were imputed with the 1000 Genomes phase 3 reference panel only[37] (Table S1).

Polygenic Risk Score

The PRS was calculated as described previously[5]. The three PRSs (for overall BC, ER-positive, and ER-negative BC) were calculated for all included individuals. The variants and their corresponding weights used in the PRS as published previously[5] and the imputation quality are listed in Table S1. The PRS for each individual was standardised to the mean from all population controls in this study and to the SD in the Breast Cancer Association Consortium (BCAC) population controls that were included in the validation data set[5]. These SDs were 0.6093, 0.6520, and 0.5920 for the overall BC PRS, ER-positive BC PRS, and ER-negative BC PRS, respectively. Using these SDs, the OR estimates for the associations of the standardised PRS_{313} in our study are directly comparable with the OR estimates reported in the BCAC population-based study[5].

Pedigree collection

Pedigrees were collected for all families and were drawn previously in the clinical genetic centres during counselling and DNA diagnostic testing of *BRCA1/2* PVs. The pedigrees were used as they were drawn in the clinic, including at least all known first- and second-

degree relatives of the genotyped individuals. Imputation of missing data is described in the supplementary material.

Family history score

A model-based family history score for BC, also called the 'polygenic load', was derived from the BOADICEA version 3 model based on the available pedigree, as described previously[7]. The polygenic load in BOADICEA is a latent polygenetic component representing the combined effect of a large number of variants each of small effect to capture the residual familial aggregation of BC and is, therefore, a measure of the BC family history[7, 10]; henceforth referred to as $BOADICEA_{FH}$. For controls with no available pedigree, $BOADICEA_{FH}$ was imputed based on the distribution of $BOADICEA_{FH}$ (normally distributed with mean=0 and SD=1).

Breast cancer lifetime risk

As all cases had developed BC, lifetime risks for developing a first breast tumour were calculated for all included individuals with the BOADICEA model[16], simulating an individual to be aged one year and unaffected. Initial lifetime risks ($BOADICEA_{ILR}$) were calculated based on *BRCA* status (all negative), pedigree information (for cases) as described above, and birth year. For individuals on whom information regarding PVs in the BC genes *CHEK2, PALB2*, and *ATM* was available, initial risks included the PV carrier status of these genes as well. The initial lifetime risks were compared with the lifetime risks calculated with the above information and the PRS_{313} ($BOADICEA_{PRS313}$).

Statistical analysis

The BC lifetime risks for cases and controls with ($BOADICEA_{PRS313}$) and without ($BOADICEA_{ILR}$) inclusion of the PRS_{313} were compared to define the change in risk category and thus advice for BC surveillance according to three different guidelines, NICE[24], NCCN[23] and IKNL[21].

To define how much of the variance in the PRS_{313} is explained by family history in this study, the degree of correlation between the standardised PRS_{313} and the $BOADICEA_{FH}$ for cases was determined by the Pearson correlation coefficient. This coefficient was calculated as well to estimate the linear correlation between the PRS_{313} of the proband (i.e. youngest BC diagnosis) and the PRS_{313} of other affected family members. If more than two family members were included, the average PRS_{313} of the family members was used. The association between overall BC (first breast tumour, invasive or *in* situ) and the PRS_{313} was determined with logistic regression using generalised estimating equations (GEE), adjusting for age and family history ($BOADICEA_{FH}$). Standard errors were corrected to account for relatedness of individuals using a robust estimator of the variance. To

reduce overfitting, association analyses included only cases that were not part of the development dataset for the PRS_{313} as described in Mavaddat et al.[5]

In a secondary analysis, we determined the association of the PRS_{313} with invasive and *in situ* BC risk separately. Cases that developed an invasive BC after the development of an *in situ* BC were only included in the invasive BC analysis with the age of diagnosis of the invasive breast tumour. Two of these cases were excluded because their age of diagnosis of invasive breast tumour was unknown.

In addition, the association between BC risk and the prevalence of a truncating variant in each of the 34 genes included in the BRIDGES gene panel[29] was determined with a two-sided Fisher Exact test.

Statistical significance was established at 5%. Analysis was performed using R version 4.0.3[38].

Results

The analyses included 3,918 cases from 3,492 families and 3,474 female population controls. In the association analyses, a subset of cases were included, i.e., those not included previously in the development dataset of the PRS_{313}[5]. These comprised 1,968 cases from 1,602 families (Figure S1, Table 1).

Characteristics of the included cases and controls are shown in Table 1. The mean age at last follow up for controls and age at diagnosis for cases was similar, 45 years, with an age range between 18 and 93 years. Most of the included cases had an invasive breast tumour (91%), 8% an *in situ* breast tumour and 1% a tumour of unknown invasiveness. Of all included cases, 18% developed a second breast tumour. The standardised PRS_{313} was higher for cases compared with controls with a mean of 0.71 (SD=0.96) compared with 0 for controls (SD=1.03). Distribution curves and descriptives of the standardised PRS_{313}, ER-positive PRS_{313}, and ER-negative PRS_{313} are shown in Figures S2 and S3 and Tables S2 and S3. In total, 218 (8.4%) cases and 47 (1.8%) controls were carriers of a truncating PV in either *ATM, CHEK2* or *PALB2,* excluding PVs in the last exon.

Table 1. Characteristics of participants

		Population controls	Family-based cases	Family-based cases – subset[a]
N		3,474	3,918	1,968
Families			3,492	1,602
Relatives per family included	**1**	3,474	3,099	1,263
	2	0	364	309
	3	0	25	25
	4	0	4	3
Study	**ABCS**	1,563	904	82
	HEBON	0	2,248	1,671
	ORIGO	987	0	0
	RBCS	924	766	215
Array	**GSA**		1,781	1,781
	iCOGS	2,388	1,680	163
	OncoArray	1,086	457	24
Age	**Mean**	45,6	45,1	46,8
	Range	18-93	21-91	21-91
First breast cancer	**Invasive**	NA	3,575	1,630
	In situ	NA	312	308
	Unknown	NA	31	30
ER status	**Positive**	NA	1,755	927
	Negative	NA	488	213
	Unknown	NA	1,675	828
Second breast tumour (N)		NA	719	327
Age	**Mean**	NA	52.6	52.9
	Range	NA	26-80	26-79
	Unknown	NA	130	29
Invasiveness	**Invasive**	NA	460	220
	In situ	NA	116	77
	Unknown	NA	144	30
ER status	**Positive**	NA	290	153
	Negative	NA	49	21
	Unknown	NA	380	153
Gene panel results	**All**	2,584	2,586	1,586
	No PV	2,537	2,369	1,463
	CHEK2 PV	31	167	98
	ATM PV	9	39	18
	CHEK2+ATM PV	0	2	1
	PALB2 PV	7	10	6
Standardised PRS_{313} (SD)	**Overall BC**	0 (1.03)	0.71 (0.96)	0.64 (0.88)
	ER+ BC	0 (1.03)	0.72 (0.97)	0.65 (0.88)
	ER- BC	0 (1.01)	0.45 (0.94)	0.29 (0.85)
$BOADICEA_{FH}$	**Mean (SD)**	0 (0.99)	0.55 (0.39)	0.69 (0.35)
Affected FDR	**0**	NA	1,125	
	1	NA	1,454	
	2	NA	555	
	>2	NA	176	
Affected SDR	**0**	NA	1,360	
	1	NA	1086	
	2	NA	583	
	>2	NA	281	
	Unknown	NA	615	

[a]Cases included in the association analyses which were not part of the development dataset for the PRS_{313} as described in Mavaddat et al.[5]
Abbreviations: $BOADICEA_{FH}$, Polygenic Load in calculated in the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; FDR, First Degree Relatives; N, Number of individuals; PRS, Polygenic Risk Score; PV, Pathogenic Variant; SD, Standard Deviation; SDR, Second Degree Relatives

Gene panel results

The BRIDGES study[29] completed sequencing for 2,037 cases with clinical data and 2,584 controls. Truncating (likely) PVs were found in 22 of 34 genes for 227 (11.1%) cases and 105 (4.1%) controls (Table S4). The majority (6.4% of the cases; 1.2% of the controls) had a truncating variant in *CHEK2,* nearly all the founder PV c.1100delC. In addition, truncating variants were relatively frequently found in *ATM, FANCM* and *PALB2* (1.8%, 0.7%, 0.6% of the cases and 0.3%, 0.6% and 0.3% of the controls respectively). The number of (pathogenic) missense variants are listed in Table S5.

PRS-based individualised risk score

Adding the PRS_{313} into the BOADICEA model ($BOADICEA_{PRS313}$) changed the absolute lifetime risk for almost all women (Figure 1), up to 34.5% for cases and up to 22.1% for controls (Figure S4, and Table S6). Clinically relevant shifts, i.e. from one to another screening category, as based on the IKNL[21], NICE[24], or NCCN[23] guidelines, were 32.4%, 36.0%, and 25.7% respectively for 1,331 cases without a gene test-result (i.e. only tested negative for a BRCA1/2 PV in diagnostic setting) (Tables 2, S7, S8). Similar results were seen for 2,369 cases that were known non-carrier of a PV in *PALB2*, *CHEK2* and *ATM*. In both groups and all age categories, a higher percentage of cases shifted to the moderate and high-risk category compared to the low-risk category (Table S9). Change towards higher risk categories was less frequent in controls than in cases (Tables S7 and S8). For cases carrying a PV in *ATM* or *CHEK2*, the proportions changing risk category were 26.3% and 17.9%, respectively, for IKNL, and 23.4% and 17.9% for NICE guidelines, but substantially lower based on the NCCN guideline (6.7% and 0.0%); this was due to the single cut-off point of 20% in the NCCN guideline. The 10 *PALB2* PV carriers in the study did not change risk category for either three guidelines.

Of the 890 controls without a gene-test result for *ATM, CHEK2*, or *PALB2* status, 4.4%, 12.0%, and 4.4% changed to another risk category based on the IKNL, NICE, and NCCN, guidelines respectively. Similar results were seen for the group where no PV was found. For *CHEK2* PV carriers, and to a lesser extent *ATM* PV carriers, these percentages were higher. Similar to cases, no change in risk category was seen for the 7 controls with a *PALB2* PV, carriers with either of three guidelines.

The distributions of the absolute lifetime risk after including the PRS_{313} for all groups ($BOADICEA_{prs313}$) are shown in Figure S5.

Table 2. Breast cancer lifetime risk category change based on the IKNL guideline

Group	BOADICEA Lifetime risk		No gene-test result		Non-PV carriers		*CHEK2* PV carriers[a]		*ATM* PV carriers[a]		*PALB2* PV carriers	
	without PRS_{313}	Including PRS_{313}	N	% change	N	% change	N	% change	N	% change	N	% change
Cases	**<20%**	**<20%**	697	30.4	1,126	30.1	3	70.0	NA		NA	
		>20%	305		486		7					
		20-30%	161	42.5	376	43.5	27	52.6	0	100.0	NA	
	20-30%	**<20%**	37		149		4		0			
		>30%	82		141		26		5			
	>30%	**>30%**	42	14.3	65	28.6	93	7.0	32	5.9	10	0.0
		<30%	7		26		7		2		0	
		Overall change		**32.4**		**33.9**		**26.3**		**17.9**		**0.0**
Controls	**<20%**	**<20%**	851	4.4	2,429	4.7	NA		NA		NA	
		>20%	39		118							
		20-30%	NA		NA		13	58.1	4	55.6	NA	
	20-30%	**<20%**					12		1			
		>30%					6		4			
	>30%	**>30%**	NA		NA		NA		NA		7	0.0
		<30%									0	
		Overall change		**4.4**		**4.7**		**58.1**		**55.6**		**0.0**

[a]Two individuals with both a pathogenic variant in *CHEK2* and *ATM* were excluded.

In total, 1,331 cases and 890 controls were included without a gene-test result; 2,369 cases and 2,537 controls in the non-PV carrier group; 167 cases and 31 controls in the CHEK2 PV carrier group; 39 cases and 9 controls in the ATM carrier group; 10 cases and 7 controls in the PALB2 PV carrier group.

Abbreviations: BOADICEA, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; IKNL, Netherlands Comprehensive Cancer Organisation guideline; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

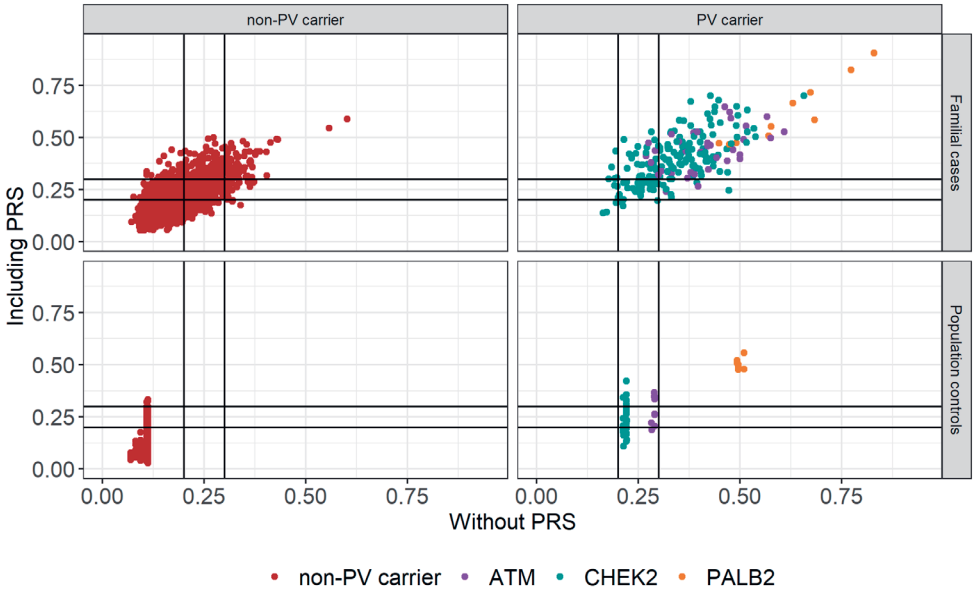


Figure 1. Change in individual breast cancer lifetime risk after including the PRS_{313}
Scatter plot of the change in breast cancer lifetime risk. For every individual, $BOADICEA_{ILR}$ was plotted against $BOADICEA_{PRS313}$. Non-carriers do not have a pathogenic variant in *ATM, CHEK2* or *PALB2* in addition to *BRCA1/2*. The solid lines represent the 20% and 30% breast cancer lifetime risk cut-off levels based on the Dutch IKNL breast cancer screening guideline[21].
Abbreviations: $BOADICEA_{ILR}$, initial breast cancer lifetime risk at age 80, based on *BRCA* status (all negative), *CHEK2*, *ATM* and *PALB2* status (if applicable), pedigree information (for cases), and birth year. $BOADICEA_{PRS313}$, breast cancer lifetime risk at age 80 including the PRS_{313} in addition to initial breast cancer lifetime risk; PRS, Polygenic Risk Score.

Table 3: Results of the association analyses between breast cancer and the PRS_{313}

		N (cases)	OR	95% CI	P-value
Main analysis	**Overall breast cancer**	1,968	1.97	1.84-2.11	$<2.00x10^{-16}$
Secondary analyses[a]	**Invasive breast cancer**	1,701	2.00	1.86-2.15	$<2.00x10^{-16}$
	***In situ* breast cancer**	262	1.69	1.50-1.89	$<2.00x10^{-16}$
Categorical PRS_{313}[b]	**0-10**	21	0.10	0.06-0.17	$<2.00x10^{-16}$
	10-20	58	0.30	0.21-0.42	$2.30x10^{-11}$
	20-40	222	0.66	0.52-0.82	$2.20x10^{-04}$
	40-60 [reference]	354	1.00	NA	NA
	60-80	491	1.37	1.13-1.66	$1.10x10^{-3}$
	80-90	396	2.27	1.84-2.79	$1.10x10^{-14}$
	90-100	426	2.29	1.86-2.83	$8.90x10^{-15}$

[a]Individuals with unknown invasiveness (N=3) and individuals with unknown age of diagnosis of the (second) invasive breast tumour (N=2) were excluded.
[b]Catagory boundaries of the PRS_{313} were -3.93; -1.27; -0.88; -0.26; 0.23; 0.84; 1.34; 3.41.
Abbreviations: CI, Confidence Interval; N, Number; OR, Odds Ratio; PRS, Polygenic Risk Score.

Correlation analysis

For cases, there was a very weak correlation between the PRS_{313} and the $BOADICEA_{FH}$ (r=0.053, p-value=$8.23x10^{-4}$); only 0.3% of the variance in the PRS_{313} is explained by family history. This poor correlation is visualised in Figures S6 and S7, where respectively the continuous and categorical $BOADICEA_{FH}$ are shown versus the PRS_{313}.

In contrast, there was a significant correlation between the PRS_{313} of the 393 probands and that of their affected family members (r=0.333, p-value= $1.00x10^{-11}$; Figure 2)

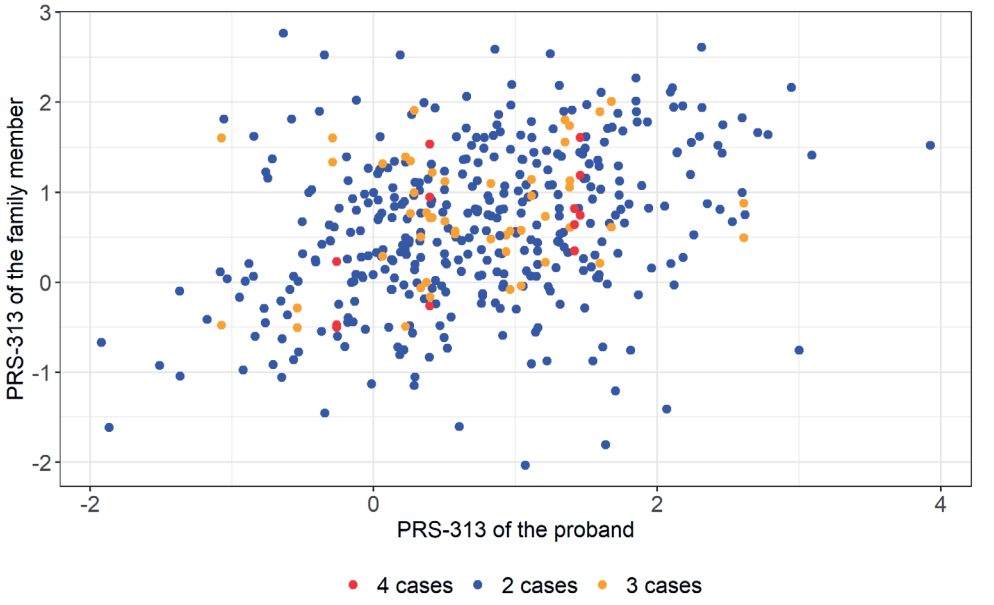


Figure 2. Correlation between the PRS_{313} of the proband and their family members
Scatter plot of the PRS_{313} of the proband (youngest breast cancer diagnosis) and their family members. Families with two individuals included are shown as blue dots, three individuals included with orange dots and four individuals included with red dots.

Abbreviations: PRS, Polygenic Risk Score.

Association analyses of PRS and breast cancer

The PRS_{313} was significantly associated with overall BC, OR per SD=1.97, 95%CI [1.84-2.11], p-value ≤$2.00x10^{-16}$ (Table 3, Figure S8). The analyses per decile followed the trend for the continuous PRS_{313}, despite that the confidence intervals of the two lowest and the highest categories did not overlap with the continuous line (Table 3; Figure S9).

Secondary analyses for invasive BC showed similar results. *In situ* BC was also significantly associated with the PRS_{313}, OR=1.69, 95%CI [1.50-1.89], p-value ≤$2.00x10^{-16}$ (Table 3, Figure S8).

Discussion

In this study, we have shown that the best performing PRS for BC at this moment[5] leads to substantially different patient stratification than the currently used in a familial cancer setting, which supports the implementation of the PRS_{313} in standard care for individuals from these families in clinical genetic services. Using a validated, comprehensive risk prediction model, BOADICEA[16, 39], pedigree-based family history can be easily combined with the individual PRS_{313}, as well as with gene panel results, to calculate a personal BC lifetime risk. We have shown that this procedure leads to a different risk category and corresponding clinical advice for substantial numbers of both non-carriers and carriers of a PV in a moderate BC risk gene. Furthermore, our results confirm the association between BC risk and the PRS_{313} in familial BC cases in the Dutch population[5, 40].

For *ATM* and *CHEK2* PV carriers, previous studies showed that including the PRS is of additive value for risk prediction and risk management[13, 14, 41]. A population-based study using a PRS of 105 variants[13] and a case-control study using a PRS of 86 variants[14] found similar results for *CHEK2* PV carriers and showed that there is no need for intensified breast screening for about 30% of these women. Dissimilar percentages were found for *ATM* carriers; about 50% based on the PRS-105, but a substantially lower percentage using the PRS-86 would not need intensified screening after including the PRS[13, 14]. These results were based on the NCCN guideline with a single cut-off of 20% guiding clinical management. Compared to these results and using the same guideline, we found a slightly higher percentage of *CHEK2* carriers in the unaffected population would have received different screening advice (39%), but a much lower percentage (7%) for cases with a positive family history. Although we did not see a shift in screening category for *PALB2* carriers, there was an absolute risk difference with a maximum of 9.8% for cases and 4.8% for population controls, corresponding to a lifetime risk range of 47%-91% for cases and 48%-56% for controls. A previous study found a similar effect for cases by including the PRS[42]. Such differences in risk could inform choices regarding preventive surgeries.

Our study did not have enough power to perform an association analysis between the PRS and BC for PV carriers in *PALB2, CHEK2* or *ATM*. However, previous studies showed that the per-SD effect size of a PRS with BC in PV carriers of moderate BC genes, such as *CHEK2*, is similar as in non-carriers or untested individuals[13, 43] but lower in carriers of PV in *BRCA1/2*[12]. Few studies have been performed on *ATM* or *PALB2* carriers, but a recent study showed that the effect sizes of the associations were in between those for *BRCA1/2* and CHEK2[14]. However, BOADICEA assumes that the effect of the PRS is similar for non-PV carriers and carriers of a PV in the genes *PALB2, ATM,* and *CHEK2,* i.e., pathogenic variants and the PRS contribute to risk independently. This may need some adjustment once the exact per SD effect sizes and interactions are known for these specific genes.

We found a higher effect size for the association between BC and the PRS_{313} (OR=1.97, 95%CI=1.84-2.11) than found in the population-based cohorts of BCAC (OR=1.61, 95%CI=1.57-1.65)[5] or the Dutch population (HR=1.56, 95%CI=1.40-1.73)[40]. This can possibly be explained by a higher genetic predisposition in families that visit the clinical genetic centre for counselling. Although we adjusted for family history, the weak correlation between the PRS and family history showed that adjustment for family history does not suffice to correct for the higher genetic predisposition based on the common low-risk variants. Furthermore, family history ($BOADICEA_{FH}$) for controls was imputed based on the assumption that the family history in controls was normally distributed with mean=0. This might have introduced a bias since the real family history of each control is unknown.

The virtually absent correlation between family history and the PRS_{313} was found in previous studies as well[7, 10, 18], underscoring the additive value of including the PRS in family-based risk prediction. However, to avoid double counting this requires careful joint consideration of family history and an explicelty measured PRS as provided by the BOADICEA algorithm. Altogether, the risk stratification by using the PRS in addition to family-based risk prediction in non-carriers and PV carriers highlights the need for using a comprehensive model including the PRS to calculate individual BC lifetime risks to guide screening and prevention advice. Of note, there is also no evidence that the per-SD PRS_{313} odds ratio differs across strata defined by lifestyle and hormonal risk factors[44].

Strengths of this study include the detailed family history that was available for cases. As we used only cases who visited clinical genetic centres for counselling, this cohort is a good representation of the families that are seen in a clinical genetic context. Furthermore, our results are based on a well-validated comprehensive risk prediction model, BOADICEA that has been shown to have accurate risk predictions for the general population and in familial setting[39, 40]

A limitation of this study is that we had only data for women of European ancestry, even though some studies have shown that (a subset of) the PRS_{313} is associated with BC in other ancestries as well[45, 46]. For Asian[45] and Latina[46] populations the PRS showed similar performance as in the European population, but for the African population[47] there was an attenuated effect size. Therefore, caution is needed for comprehensive risk prediction including the PRS for women of African ancestry.

In summary, including the PRS_{313} in family history-based risk prediction may change screening recommendations in up to 34% of the individuals from families with no PVs in any of the five BC genes modelled in BOADICEA. Adding the PRS_{313} also had a large impact on screening recommendations for *ATM* and *CHEK2* PV carriers. Because BOADICEA has been prospectively validated and calibrated[39, 40], clinical implementation of comprehensive

risk prediction should be considered, although this will be a logistic challenge for clinical genetic centres and would require clinical geneticists to become aware of its limitations.

Acknowledgements and funding

This work was supported by the Dutch Cancer Society (KWF), grant UL2014-7473.

We would like to thank Mary Velthuizen (University Medical Center Utrecht), J. Margriet Collée (Erasmus Medical Center Rotterdam), Wendy Prager-van den Smissen (Erasmus Medical Center Rotterdam), Daoud Ait Moha (The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital), Jan C. Oosterwijk (University Medical Center Groningen), Janet R. Vos (Radboud University Medical Center), Jeroen F.J. Laros (Leiden University Medical Center) and Adri J.M. Krol (Leiden University Medical Center) for collecting and converting the pedigree files of the included families.

Disclosure of potential conflicts of interest

AL is listed as an inventor of BOADICEA V5, which is commercialised through Cambridge Enterprise, part of Cambridge University.

References

1. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European journal of cancer (Oxford, England : 1990)*. Nov 2018;103:356-387. doi:10.1016/j.ejca.2018.07.005
2. Ripping TM, Verbeek AL, Fracheboud J, de Koning HJ, van Ravesteyn NT, Broeders MJ. Overdiagnosis by mammographic screening for breast cancer studied in birth cohorts in The Netherlands. *International journal of cancer*. Aug 15 2015;137(4):921-9. doi:10.1002/ijc.29452
3. 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
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. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
6. 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]
7. Lakeman IMM, Hilbers FS, Rodriguez-Girondo M, et al. 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. *Journal of medical genetics*. Sep 2019;56(9):581-589. doi:10.1136/jmedgenet-2019-106072
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. 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
10. 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
11. 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
12. Barnes D, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in Medicine*. 2020;accepted for publication

13. Gao C, Polley EC, Hart SN, et al. Risk of Breast Cancer Among Carriers of Pathogenic Variants in Breast Cancer Predisposition Genes Varies by Polygenic Risk Score. *Journal of Clinical Oncology*. 0(0):JCO.20.01992. doi:10.1200/jco.20.01992
14. Gallagher S, Hughes E, Wagner S, et al. Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. *JAMA Network Open*. 2020;3(7):e208501-e208501. doi:10.1001/jamanetworkopen.2020.8501
15. 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
16. 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
17. 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
18. 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]
19. 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
20. 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
21. IKNL. Richtlijn Borstkanker - Screening buiten het bevolkingsonderzoek. Accessed 03-12-2021, https://richtlijnendatabase.nl/richtlijn/borstkanker/screening/screening_buiten_het_bob/screening_buiten_het_bevolkingsonderzoek.html
22. Carver T, Hartley S, Lee A, et al. CanRisk Tool—A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer Epidemiology Biomarkers & Prevention*. 2021;30(3):469-473. doi:10.1158/1055-9965.Epi-20-1319
23. 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
24. 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;

25. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ (Clinical research ed)*. Oct 20 2007;335(7624):806-8. doi:10.1136/bmj.39335.541782.AD
26. van Asperen CJ, Brohet RM, Meijers-Heijboer EJ, et al. Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. *Journal of medical genetics*. Sep 2005;42(9):711-9. doi:10.1136/jmg.2004.028829
27. Schmidt MK, Hogervorst F, van Hien R, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Aug 10 2016;34(23):2750-60. doi:10.1200/jco.2016.66.5844
28. Liu J, Prager-van der Smissen WJ, Schmidt MK, et al. Recurrent HOXB13 mutations in the Dutch population do not associate with increased breast cancer risk. *Sci Rep*. Jul 18 2016;6:30026. doi:10.1038/srep30026
29. Dorling L, Carvalho S, Allen J, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *The New England journal of medicine*. Jan 20 2021;doi:10.1056/NEJMoa1913948
30. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biology*. 2016/06/06 2016;17(1):122. doi:10.1186/s13059-016-0974-4
31. Boonen R, Rodrigue A, Stoepker C, et al. Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. *Nat Commun*. Nov 22 2019;10(1):5296. doi:10.1038/s41467-019-13194-2
32. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. Jan 4 2018;46(D1):D1062-d1067. doi:10.1093/nar/gkx1153
33. 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]
34. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nature genetics*. Oct 2016;48(10):1284-1287. doi:10.1038/ng.3656
35. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*. Oct 2016;48(10):1279-83. doi:10.1038/ng.3643
36. 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
37. 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
38. R_Core_Team_(2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
39. Pal Choudhury P, Brook MN, Hurson AN, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. *Breast Cancer Research*. 2021/02/15 2021;23(1):22. doi:10.1186/s13058-021-01399-7

40. Lakeman IMM, Rodríguez-Girondo M, Lee A, et al. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. *Genetics in medicine : official journal of the American College of Medical Genetics*. Nov 2020;22(11):1803-1811. doi:10.1038/s41436-020-0884-4
41. Borde J, Ernst C, Wappenschmidt B, et al. Performance of Breast Cancer Polygenic Risk Scores in 760 Female CHEK2 Germline Mutation Carriers. *Journal of the National Cancer Institute*. Jul 1 2021;113(7):893-899. doi:10.1093/jnci/djaa203
42. Mars N, Widén E, Kerminen S, et al. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun*. Dec 14 2020;11(1):6383. doi:10.1038/s41467-020-19966-5
43. Muranen TA, Greco D, Blomqvist C, et al. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. *Genetics in medicine : official journal of the American College of Medical Genetics*. Oct 06 2016;doi:10.1038/gim.2016.147
44. Kapoor PM, Mavaddat N, Choudhury PP, et al. Combined Associations of a Polygenic Risk Score and Classical Risk Factors With Breast Cancer Risk. *Journal of the National Cancer Institute*. Mar 1 2021;113(3):329-337. doi:10.1093/jnci/djaa056
45. Ho WK, Tan MM, Mavaddat N, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun*. Jul 31 2020;11(1):3833. doi:10.1038/s41467-020-17680-w
46. Shieh Y, Fejerman L, Lott PC, et al. A polygenic risk score for breast cancer in U.S. Latinas and Latin-American women. *Journal of the National Cancer Institute*. Sep 25 2019;doi:10.1093/jnci/djz174
47. Du Z, Gao G, Adedokun B, et al. Evaluating Polygenic Risk Scores for Breast Cancer in Women of African Ancestry. *Journal of the National Cancer Institute*. Mar 26 2021;doi:10.1093/jnci/djab050

Supplementary methods

Study cohorts

HEBON

The HEBON study[1] (initiated in 1999) is an ongoing nationwide retrospective cohort study among breast cancer families with prospective follow up. Participants were invited after visiting one of the Clinical Genetic Centers in the Netherlands for breast and/or ovarian cancer counselling. Participants were asked to fill in a questionnaire about lifestyle, family history and risk factors for breast cancer. Linkage with the nationwide cancer and pathology registries is possible for follow up.

Additional selection criteria for HEBON participants included:

- At least two breast cancer cases in a family with available DNA samples
- Breast cancer diagnosis below the age of 60 years and a positive family history:
    - One first degree family member with breast cancer diagnosis below the age of 50 OR
    - Two first or second-degree family members with breast cancer diagnosis below the age of 60

ABCS-F and RBCS

The ABCS-F[2] and RBCS[3] case-cohorts included also breast cancer cases who visited the Clinical Genetic Centres of the Netherlands Cancer Institute in Amsterdam or the Erasmus Medical Center in Rotterdam, respectively. No additional selection criteria were used for ABCS-F and RBCS cases. 151 individuals from the ABCS-F study and 469 individuals from the RBCS study are included in the HEBON study as well and shown as HEBON cases in Table 1.

Quality control procedure

For the 2,179 breast cancer cases without a *BRCA1/2* pathogenic variant that were genotyped with the GSA array, quality control was performed with Plink version 1.9, which excluded 8,408 SNPs with a call rate below 95%. Another 712 SNPs were removed because of a deviation from Hardy-Weinberg equilibrium in controls at $P<1x10^{-12}$. In total, 124 individuals were excluded of which 62 individuals with a call rate below 95%, 7 individuals because they were genotypically not female or the gender was uncertain, and 17 individuals because of a sample swab. After population stratification analysis, 28 individuals were excluded because of non-European genotype (>3 SD).

Imputation pedigrees

In total, 3,492 pedigrees were collected for this study. These pedigrees consisted of 202,680 individuals (49% female) of which 12,785 individuals were affected with breast cancer.

If the age of breast cancer diagnosis for a family member was not known (n=1,272), a conditional average age was estimated given the age at last follow up of the individual and the breast cancer incidence in the Netherlands. Furthermore, for all affected individuals with breast cancer, ovarian cancer, prostate cancer or pancreatic cancer the year of birth was imputed, if this was not yet available, based on the year of birth of the closest relative (25 year difference for parents and children, average for siblings). If the age of last follow up was not known, this age was calculated based on the date of the last update of a pedigree and the year of birth.

Supplementary figures and tables

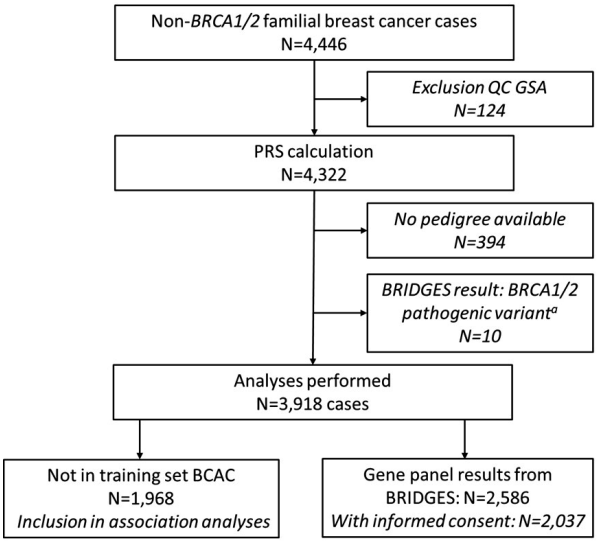


Figure S1: Flow scheme of the selection procedure
Breast cancer cases were selected from the ABCS, HEBON and RBCS studies. Details of the quality control procedure are described above. Absolute lifetime risks were calculated for all included cases (N=3,918). To exclude overlap of cases with the development dataset for the PRS_{313} [4], only 1,968 cases were included in the association analyses. For the majority of cases gene panel information was available. For cases of whom we did not have informed consent to report the clinical relevant results, only pseudoanonymized information about pathogenic variants in *ATM, CHEK2*, and *PALB2* was available (N=549). For the cases with informed consent, the number of pathogenic variants and missense variants are shown in Table S3.
[a]carriers of a pathogenic variant or family member of a carrier of a pathogenic variant in *BRCA1* or *BRCA2*.
Abbreviations: BCAC, Breast Cancer Association Consortium; BRIDGES, Breast cancer Risk after Diagnostic GEne Sequencing; PRS, Polygenic Risk Score.

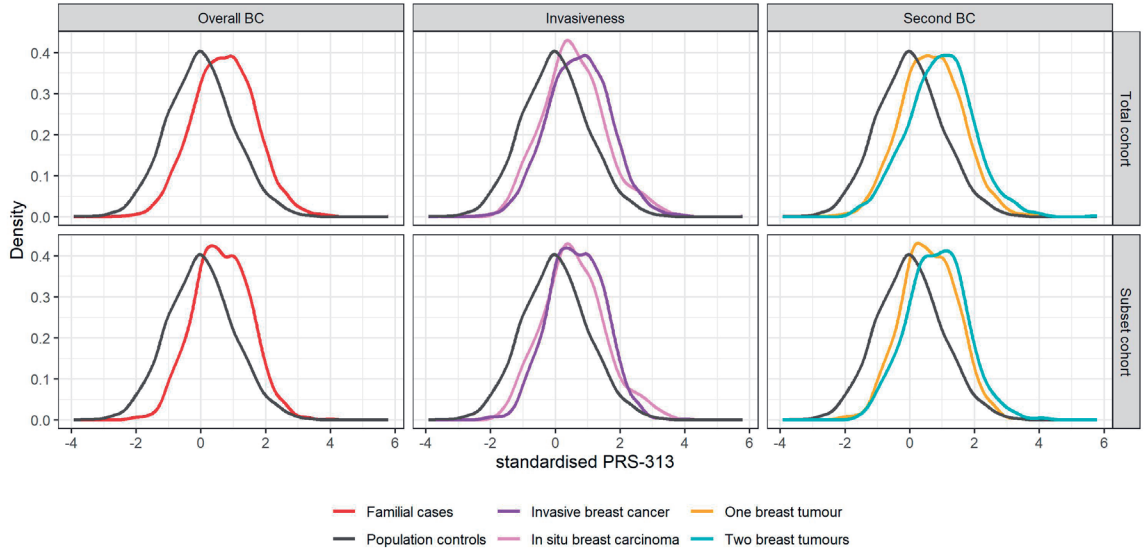


Figure S2: Density curves of the PRS_{313}
Distribution of the PRS_{313} in the included 3,474 population controls (grey line) and 3,918 and 1,968 breast cancer cases (red line) in the total and subset cohort respectively. For the invasiveness figure, 3 cases were excluded for which invasiveness for the first and/or second breast tumour was unknown. In the total cohort 3,653 and 262 cases were included with invasive (purple line) and in situ (pink line) breast cancer respectively. For the subset cohort this was 1,703 and 262. In the right figure, 719 and 327 breast cancer cases with a second breast tumour (blue line) were included in the total and subset cohort respectively.
Abbreviations: BC, Breast Cancer; PRS, Polygenic Risk Score.

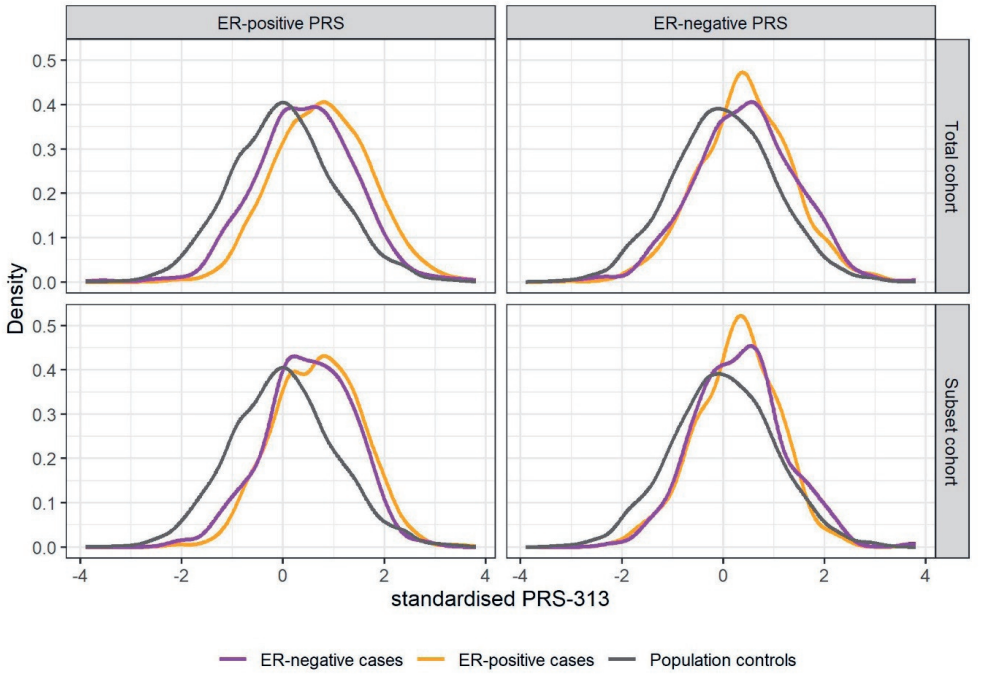


Figure S3: Density curves of the ER-positive and ER-negative PRS_{313}
Distribution of the ER-negative (left figures) and ER-positive (right figures) PRS_{313} for cases with an ER-negative (purple line) and ER-positive (orange line) first breast tumour. As a reference, the distribution of these PRS in population controls are shown as well (grey line). In the total cohort, 1,755 and 488 breast cancer cases are included with a first ER-positve and ER-negative breast tumour respectively. For the subset cohort this was 927 and 213 respectively.
Abbreviations: ER, Estrogen Receptor; PRS, Polygenic Risk Score

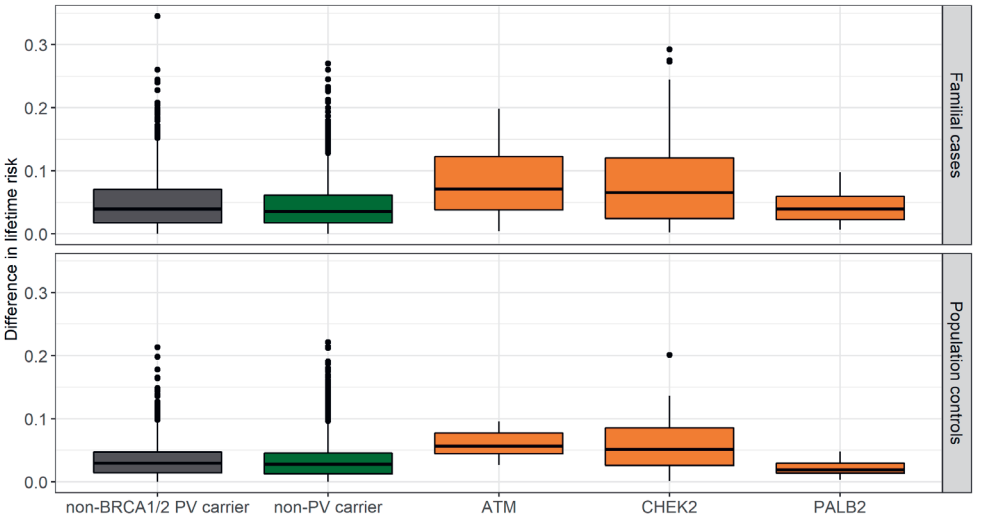


Figure S4: Difference in breast cancer lifetime risk score calculated by BOADICEA
Boxplot of the difference in breast cancer lifetime risk between the basic calculation in BOADICEA and after including the PRS_{313}. The basic calculation included birth year, gene panel results and for cases a pedigree of their family in addition. Non-carriers are the group of which we know that they do not have a pathogenic variant in *ATM, CHEK2* and *PALB2* in addition to *BRCA1/2*.
Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PV, Pathogenic Variant.

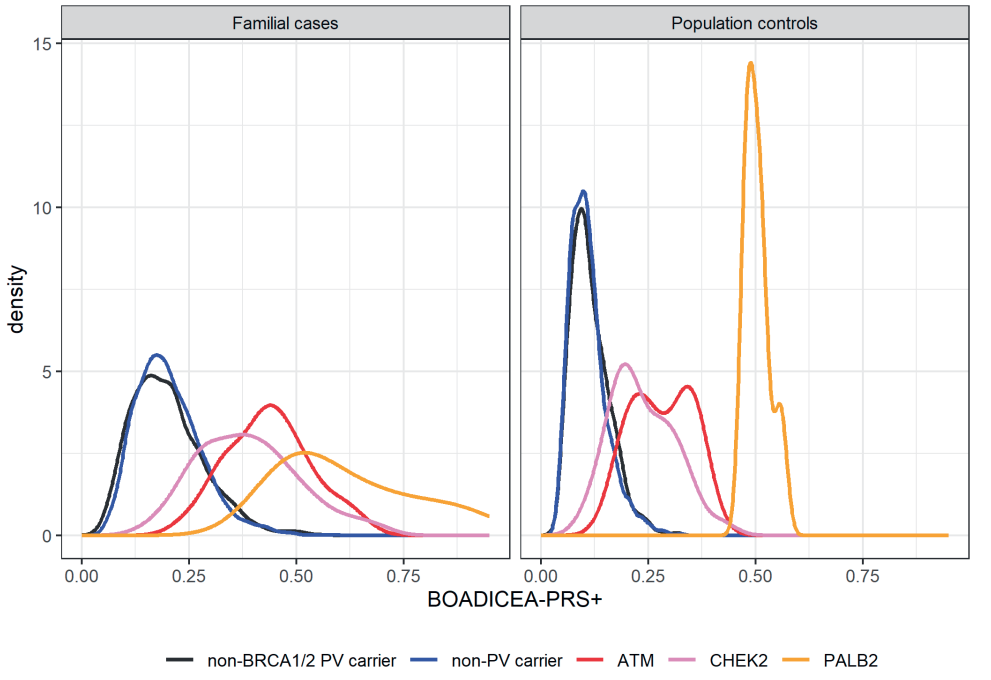


Figure S5. Distribution of breast cancer lifetime risk after including the PRS_{313}
Density plots of the distribution in breast cancer lifetime risk calculated with BOADICEA including birth cohort, gene panel results, pedigree-based family history for cases and the PRS_{313}.
Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PV, Pathogenic Variant; PRS, Polygenic Risk Score

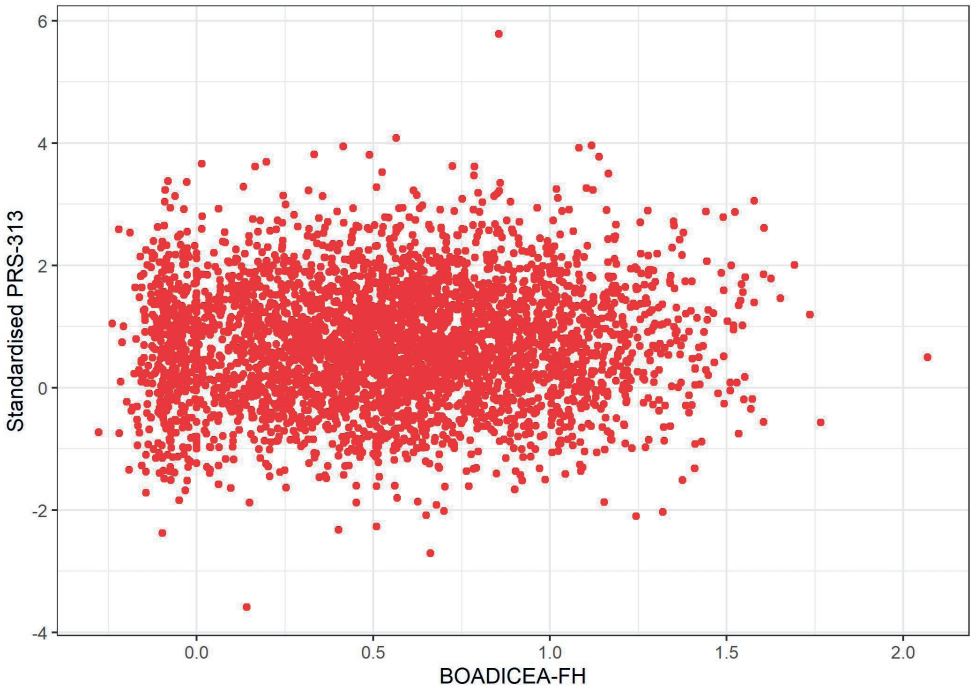


Figure S6. Correlation plot between de $BOADICEA_{FH}$ and the PRS_{313}
For all included breast cancer cases (N=3,918), the individual $BOADICEA_{FH}$ (polygenic load) is plotted against the PRS_{313}. $BOADICEA_{FH}$ was calculated with BOADICEA based on the pedigree without inclusion of the PRS_{313}.
Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; FH, Family History; PRS, Polygenic Risk Score.

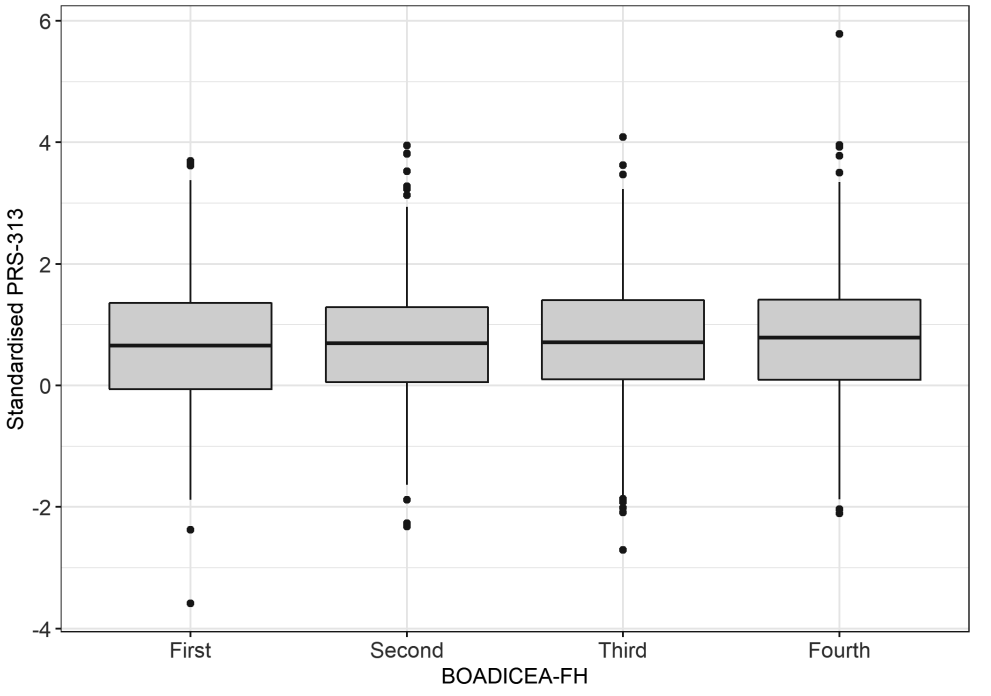


Figure S7: PRS_{313} distribution by quartiles of $BOADICEA_{FH}$
The PRS_{313} distribution for all included cases (N=3,918) separated by quartiles of the individual $BOADICEA_{FH}$ (polygenic load). $BOADICEA_{FH}$ was calculated with BOADICEA based on the pedigree without inclusion of the PRS_{313}.
Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; FH, Family History; PRS, Polygenic Risk Score.

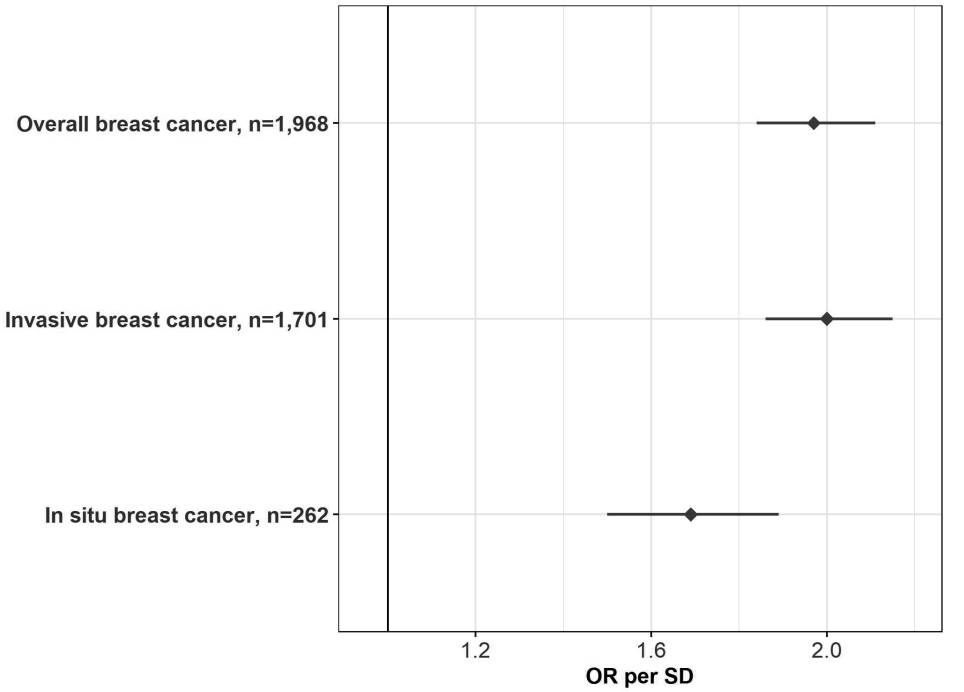


Figure S8: Association between the PRS_{313} and breast cancer
Visualisation of the effect sizes and 95% confidence intervals of the association between the PRS_{313} and breast cancer. The corresponding OR and included breast cancer cases are shown in Table 3. Abbreviations: BC, Breast Cancer; OR, Odds Ratio; PRS, Polygenic Risk Score

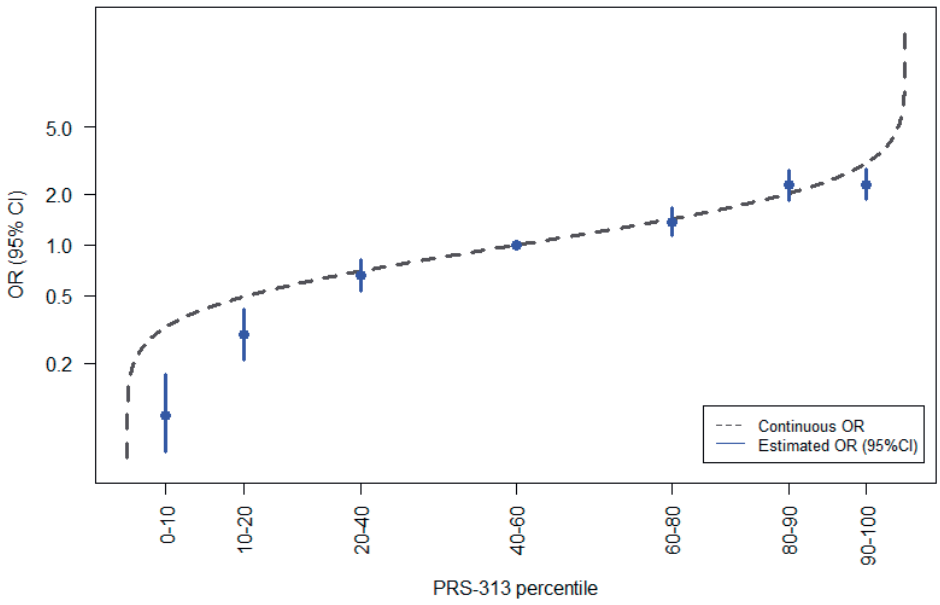


Figure S9: Association between the PRS and breast cancer by percentiles of the PRS_{313}
Plot of the effect size of the association between the continuous PRS_{313} (grey line) and breast cancer and the categorical PRS_{313} (blue dots) and breast cancer. Corresponding OR and 95% confidence intervals are shown in Table 3.
Abbreviations: CI, Confidence Interval; OR, Odds Ratio; PRS, Polygenic Risk Score.

Table S1: common low risk variants included in the PRS_{313} (large Excel file)
Available upon request / see online material. This table is partly published before by Mavaddat et al.[4] We added the imputation quality in this study.

Table S2: Descriptives of the standardised PRS_{313}

	Total cohort			Family-based cases – subset[c]		
	N	Mean PRS_{313}	SD PRS_{313}	N	Mean PRS_{313}	SD PRS_{313}
All cases	3,918	0.71	0.96	1,968	0.64	0.88
Invasive cases[a]	3,653	0.73	0.96	1,703	0.65	0.86
***In situ* only cases[b]**	262	0.56	0.96	262	0.56	0.96
1 breast tumour	3,199	0.66	0.95	1,641	0.60	0.87
2 breast tumours	719	0.95	1.01	327	0.83	0.90
Population controls	3,474	0	1.03	NA	NA	NA

[a]Invasive first or second tumour
[b]no invasive first or second tumour
[c]Cases included in the association analyses which were not part of the development dataset for the PRS_{313} as described in Mavaddat et al.[4]
Abbreviations: N, Number; NA, Not Applicable; PRS, Polygenic Risk Score

Table S3: Descriptives of the standardised ER-positive and ER-negative PRS_{313}

Group	PRS	Total cohort			Family-based cases – subset[c]		
		N	Mean PRS	SD PRS	N	Mean PRS	SD PRS
ER-positive BC	ER-positive PRS	1,755	0.78	0.92	927	0.68	0.86
ER-negative BC	ER-positive PRS	488	0.43	0.98	213	0.51	0.85
ER-positive BC	ER-negative PRS	1,755	0.76	0.93	927	0.66	0.85
ER-negative BC	ER-negative PRS	488	0.46	0.97	213	0.52	0.85

[a]Invasive first or second tumour
[b]no invasive first or second tumour
[c]Cases included in the association analyses which were not part of the development dataset for the PRS_{313} as described in Mavaddat et al.[4]
Abbreviations: N, Number; NA, Not Applicable; PRS, Polygenic Risk Score

Table S4: Truncating variants in BRIDGES gene panel

Gene	Cases, N=2,037[a]		Controls, N=2,584[a]		OR	95% CI	P-value
	N	%	N	%			
ABRAXAS1	1	0.0	0	0.0	NA	NA	NA
AKT1	0	0.0	0	0.0	NA	NA	NA
ATM	**36**	**1.8**	**9**	0.3	**5.15**	**2.42-12.18**	**1.00x10^{-06}**
BARD1	1	0.0	1	0.0	1.27	0.02-99.55	1.00
BRCA1	NA	NA	NA	NA	NA	NA	NA
BRCA2	NA	NA	NA	NA	NA	NA	NA
BRE	0	0.0	0	0.0	NA	NA	NA
BRIP1	4	0.2	5	0.2	1.01	0.20-4.72	1.00
CDH1	0	0.0	0	0.0	NA	NA	NA
CHEK2	**131**	**6.4**	**31**	**1.2**	**5.66**	**3.78-8.70**	**<2.00x10^{-16}**
c.1100delC[b]	130		30				
Other	1						
EPCAM	0	0.0	2	0.1	NA	NA	NA
FANCC	5	0.2	8	0.3	0.79	0.20-2.75	0.80
FANCM	14	0.7	16	0.6	1.11	0.50-2.44	0.90
GEN1	0	0.0	0	0.0	NA	NA	NA
MEN1	0	0.0	0	0.0	NA	NA	NA
MLH1	0	0.0	0	0.0	NA	NA	NA
MRE11A	1	0.0	3	0.1	0.42	0.01-5.27	0.60
MSH2	0	0.0	2	0.1	NA	NA	NA
MSH6	1	0.0	0	0.0	NA	NA	NA
MUTYH	3	0.1	2	0.1	1.9	0.22-22.81	0.70
NBN	2	0.1	3	0.1	0.85	0.07-7.39	1,00
NF1	2	0.1	0	0.0	NA	NA	NA
PALB2	12[c]	0.6	7	0.3	2.18	0.79-6.55	0.10
PIK3CA	0	0.0	0	0.0	NA	NA	NA
PMS2	1	0.0	2	0.1	0.63	0.01-12.19	1.00
PTEN	1	0.0	1	0.0	1.27	0.02-99.55	1.00
RAD50	4	0.2	7	0.3	0.72	0.16-2.85	0.80
RAD51C	1	0.0	0	0.0	NA	NA	NA
RAD51D	5	0.2	0	0.0	NA	NA	NA
RECQL	2	0.1	3	0.1	0.85	0.07-7.39	1.00
RINT1	0	0.0	2	0.1	NA	NA	NA
STK11	0	0.0	0	0.0	NA	NA	NA
TP53	0	0.0	0	0.0	NA	NA	NA
XRCC2	0	0.0	1	0.0	NA	NA	NA
Total	227	11.1	105	4.1	-	-	-

[a]Cases and controls were included in the analyses described by Dorling et al.[5]
[b]of which 6 homozygous in cases and 1 homozygous in controls
[c]In addition to inclusion criteria for truncating variants in BRIDGES, 4 *PALB2* truncating variants in the last exon were added.
Abbreviations: CI, Confidence Interval; N, Number; NA, Not Applicable; OR, Odds Ratio.

Table S5: Missense variants in BRIDGES gene panel

Gene	Cases; N=2,038[a]		Controls, N=2,584[a]	
	Total[b]	P/LP[c]	Total[b]	P/LP[c]
ABRAXAS1	3	NA	5	NA
AKT1	2	NA	6	NA
ATM	121	5	113	4
BARD1	25	0	26	0
BRCA1	42	NA	49	NA
BRCA2	109	NA	127	NA
BRE	0	NA	0	NA
BRIP1	34	NA	41	NA
CDH1	26	NA	28	NA
CHEK2	64	8	34	2
EPCAM	9	NA	18	NA
FANCC	28	NA	23	NA
FANCM	64	NA	62	NA
GEN1	38	NA	32	NA
MEN1	4	NA	2	NA
MLH1	19	NA	21	NA
MRE11A	16	NA	19	NA
MSH2	42	NA	56	NA
MSH6	51	NA	52	NA
MUTYH	28	NA	33	NA
NBN	35	NA	23	NA
NF1	30	NA	34	NA
PALB2	23	0	23	0
PIK3CA	6	NA	10	NA
PMS2	37	NA	28	NA
PTEN	3	NA	7	NA
RAD50	50	NA	46	NA
RAD51C	9	1	9	0
RAD51D	6	0	10	0
RECQL	16	NA	20	NA
RINT1	39	NA	47	NA
STK11	0	NA	1	NA
TP53	14	4	10	0
XRCC2	6	NA	13	NA
Total	999	18	1,028	6

[a]Cases and controls were included in the analyses described by Dorling et al.[5]
[b]Total number of missense variants detected, not corrected for individuals who carry more than one missense variant in a single gene.
[c]For genes in which pathogenic variants are associated with breast cancer[5], missense variant interpretation was performed by using the ClinVar database[6].
Abbreviations: N, Number; NA, Not Applicable; P, Pathogenic; LP, Likely Pathogenic.

Table S6: Absolute change in breast cancer lifetime risk after including the PRS_{313}

	Cases			Controls		
	Min	**Mean**	**Max**	**Min**	**Mean**	**Max**
Non-*BRCA1/2* PV carriers	0	5.0	34.5	0	3.5	21.3
Non-carriers	0	4.5	27.0	0	3.3	22.1
***ATM* PV carriers**[a]	0.4	8.0	19.8	2.6	5.9	9.6
***CHEK2* PV carriers**[a]	0.3	8.1	29.3	0.1	5.9	20.1
***PALB2* PV carriers**	0.7	4.4	9.8	0.3	2.2	4.8

[a]Two cases with both a pathogenic variant in *CHEK2* and *ATM* were excluded.
In total, 1,331 cases and 890 controls were included without a gene-test result; 2,369 cases and 2,537 controls in the non-PV carrier group; 167 cases and 31 controls in the CHEK2 PV carrier group; 39 cases and 9 controls in the ATM carrier group; 10 cases and 7 controls in the PALB2 PV carrier group.
Abbreviations: Min, Minimum; Max, Maximum; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

Table S7: Breast cancer lifetime risk category change based on the NCCN guideline

Group	BOADICEA Lifetime risk		No gene-test result		Non-PV carriers		*CHEK2* PV carriers[a]		*ATM* PV carriers[a]		*PALB2* PV carriers	
	Without PRS_{313}	Including PRS_{313}	N	% change	N	% change	N	% change	N	% change	N	% change
Cases	**<20%**	**<20%**	697	30.4	1,126	30.1	3	70.0	0	0.0	0	0.0
		>20%	305		486		7		0		0	
	>20%	**>20%**	292	11.2	605	20.1	153	2.5	39	0.0	10	0.0
		<20%	37		152		4		0		0	
		Overall change		**25.7**		**26.9**		**6.6**		**0.0**		**0.0**
Controls	**<20%**	**<20%**	851	4.4	2,419	4.7	NA		NA		NA	
		>20%	39		118							
	>20%	**>20%**	NA		NA		19	38.7	8	11.1	7	0.0
		<20%					12		1		0	
		Overall change		**4.4**		**4.7**		**38.7**		**11.1**		**0.0**

[a]Two cases with both a pathogenic variant in *CHEK2* and *ATM* were excluded.

In total, 1,331 cases and 890 controls were included without a gene-test result (no BRCA1/2 PV); 2,369 cases and 2,537 controls in the non-PV carrier group; 167 cases and 31 controls in the CHEK2 PV carrier group; 39 cases and 9 controls in the ATM carrier group; 10 cases and 7 controls in the PALB2 PV carrier group.

Abbreviations: BOADICEA, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; NCCN, the National Comprehensive Cancer Network guideline; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

Table S8: Breast cancer lifetime risk category change based on the NICE guideline

Group	BOADICEA Lifetime risk		No gene-test result		Non-PV carriers		*CHEK2* PV carriers[a]		*ATM* PV carriers[a]		*PALB2* PV carriers	
	Without PRS_{313}	Including PRS_{313}	N	% change	N	% change	N	% change	N	% change	N	% change
Cases	**<17%**	**<17%**	478	38.5	699	37.1	1	0.0	NA		NA	
		>17%	299		413		0					
		17-30%	332	34.3	799	31.5	34	48.5	0	100.0	NA	
	17-30%	**<17%**	68		203		1		0			
		>30%	105		164		31		5			
	>30%	**>30%**	42	14.3	65	28.6	93	7.0	32	5.9	10	0.0
		<30%	67		26		7		2		0	
		Overall change	**36.0**		**34.0**		**23.4**		**17.9**		**0.0**	
Controls	**<17%**	**<17%**	783	12.0	2,289	9.8	NA		NA		NA	
		>17%	107		248							
		17-30%	NA		NA		20	35.5	5	44.4	NA	
	17-30%	**<17%**					5		0			
		>30%					6		4			
	>30%	**>30%**	NA		NA		NA		NA		7	0.0
		<30%									0	
		Overall change	**12.0**		**9.8**		**35.5**		**44.4**		**0.0**	

[a]Two cases with both a pathogenic variant in *CHEK2* and *ATM* were excluded.

In total, 1,331 cases and 890 controls were included without a gene-test result; 2,369 cases and 2,537 controls in the non-PV carrier group; 167 cases and 31 controls in the CHEK2 PV carrier group; 39 cases and 9 controls in the ATM carrier group; 10 cases and 7 controls in the PALB2 PV carrier group.

Abbreviations: BOADICEA, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; NICE, the National Institute for Health and Care Excellence guideline; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

Table S9: Breast cancer lifetime risk by age of breast cancer diagnosis for cases based on the Dutch IKNL guideline

Group		<40 years		40-50 years		≥50 years	
	BOADICEA LTR	Without PRS_{313}	Including PRS_{313}	Without PRS_{313}	Including PRS_{313}	Without PRS_{313}	Including PRS_{313}
No gene-test result	**<20%**	403 (87%)	305 (66%)	377 (74%)	257 (50%)	222 (62%)	172 (48%)
	20-30%	58 (13%)	127 (27%)	111 (22%)	186(36%)	111 (31%)	122 (34%)
	>30%	1 (0%)	30 (6%)	24 (5%)	69 (13%)	24 (7%)	63 (17%)
Non-PV carriers	**<20%**	475 (81%)	367 (62%)	706 (65%)	557 (52%)	431 (61%)	354 (50%)
	20-30%	96 (16%)	183 (31%)	328 (30%)	395 (37%)	242 (34%)	267 (38%)
	>30%	17 (3%)	38 (6%)	44 (4%)	126 (12%)	30 (4%)	82 (12%)
CHEK2 PV carriers[a]	**<20%**	4 (8%)	3 (6%)	4 (5%)	1 (1%)	2 (4%)	3 (7%)
	20-30%	17 (35%)	12 (24%)	22 (30%)	11 (15%)	18 (40%)	13 (29%)
	>30%	28 (57%)	34 (69%)	47 (46%)	61 (84%)	25 (56%)	29 (64%)
ATM PV carriers[a]	**<20%**	NA	NA	NA	NA	NA	NA
	20-30%	2 (20%)	1 (10%)	2 (12%)	1 (6%)	1 (8%)	0 (0%)
	>30%	8 (80%)	9 (90%)	15 (88%)	16 (94%)	11 (92%)	12 (100%)
PALB2 PV carriers	**<20%**	NA	NA	NA	NA	NA	NA
	20-30%	NA	NA	NA	NA	NA	NA
	>30%	4 (100%)	4 (100%)	5 (100%)	5 (100%)	1 (100%)	1 (100%)

[a]Two cases with both a pathogenic variant in *CHEK2* and *ATM* were excluded.

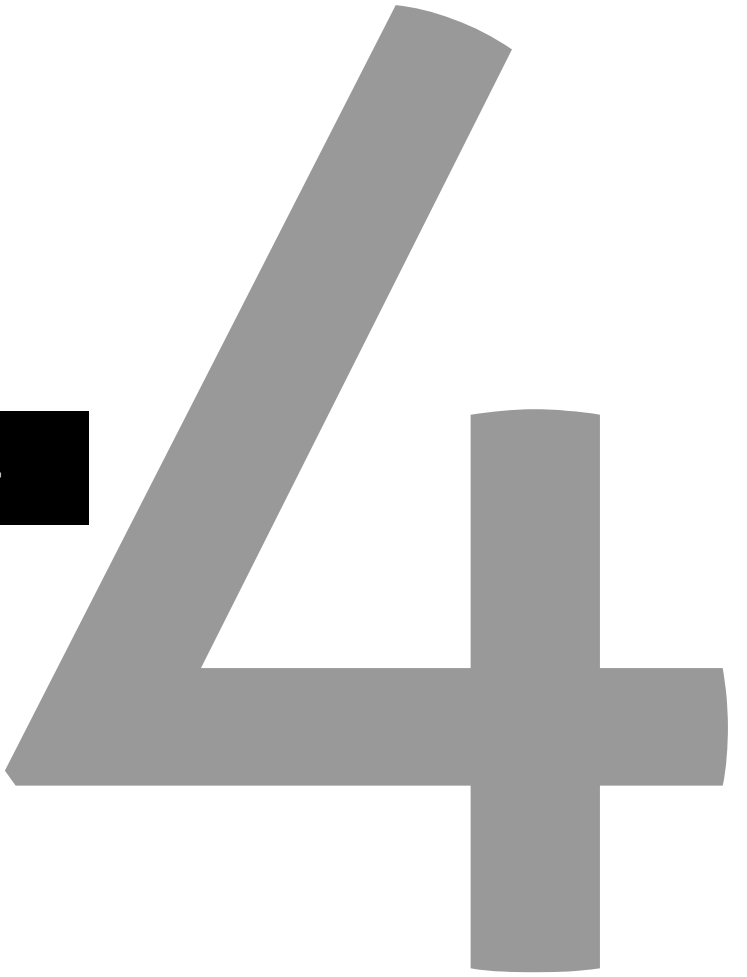
In total, 1,331 cases were included without a gene-test result; 2,369 cases in the non-PV carrier group; 167 cases in the CHEK2 PV carrier group; 39 cases in the ATM carrier group; 10 cases in the PALB2 PV carrier group.

Abbreviations: BOADICEA, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; IKNL, Netherlands Comprehensive Cancer Organisation guideline; LTR, Life Time Risk; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

Supplementary references

1. van Asperen CJ, Brohet RM, Meijers-Heijboer EJ, et al. Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. *Journal of medical genetics*. Sep 2005;42(9):711-9. doi:10.1136/jmg.2004.028829
2. Schmidt MK, Hogervorst F, van Hien R, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Aug 10 2016;34(23):2750-60. doi:10.1200/jco.2016.66.5844
3. Liu J, Prager-van der Smissen WJ, Schmidt MK, et al. Recurrent HOXB13 mutations in the Dutch population do not associate with increased breast cancer risk. *Sci Rep*. Jul 18 2016;6:30026. doi:10.1038/srep30026
4. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
5. Dorling L, Carvalho S, Allen J, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *The New England journal of medicine*. Jan 20 2021;doi:10.1056/NEJMoa1913948
6. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. Jan 4 2018;46(D1):D1062-d1067. doi:10.1093/nar/gkx1153

CHAPTER 4



Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort

Inge M.M. Lakeman, Mar Rodríguez-Girondo, Andrew Lee, Rikje Ruiter, Bruno H. Stricker, Sara R.A. Wijnant, Maryam Kavousi, Antonis C. Antoniou, Marjanka K. Schmidt, André G. Uitterlinden, Jeroen van Rooij, and Peter Devilee

Lakeman IMM, et al. Genetics in Medicine 2020;22:1803-1811

Abstract

Purpose: We evaluated the performance of the recently extended Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA version 5) in a Dutch prospective cohort, using a Polygenic Risk Score based on 313 breast cancer-associated variants (PRS_{313}), and other, non-genetic risk factors.

Methods: Since 1989, 6,522 women without breast cancer (BC) aged 45 or older of European descent were included in the Rotterdam Study. The PRS_{313} was calculated per 1 standard deviation (SD) in controls from the Breast Cancer Association Consortium (BCAC). Cox regression analysis was performed to estimate the association between the PRS_{313} and incident BC risk. Cumulative 10-year risks were calculated with BOADICEA including different sets of variables (age, risk factors and PRS_{313}). C-statistics were used to evaluate discriminative ability.

Results: In total, 320 women developed BC. The PRS_{313} was significantly associated with BC (HR per SD of 1.56, 95%CI [1.40-1.73]). Using 10-year risk estimates including age and the PRS_{313}, other risk factors improved the discriminatory ability of the BOADICEA model marginally, from a C-statistic of 0.636 to 0.653.

Conclusion: The effect-size of the PRS_{313} is highly reproducible in the Dutch population. Our results validate the BOADICEA v5 model for BC risk assessment in the Dutch general population.

Introduction

Breast cancer is the most common cancer among women in Europe[1]. In the Netherlands, the average lifetime risk for developing invasive breast cancer is 13.6% for each woman, with the incidence peaking between 60-70 years of age[2]. Mammographic screening has decreased breast cancer mortality at the cost of detecting more disease that otherwise would not have become clinically apparent[3, 4]. Based on the UK guidelines, for every 10,000 women invited for screening at age 50 for the following 20 years, 43 deaths would be prevented, while 129 breast cancers would be overdiagnosed[5]. Furthermore, breast cancer screening inevitably yields false positives which can lead to anxiety[6]. Improvement of this benefit-to-harm ratio could be achieved by targeting women who benefit the most from screening, in particular those in the highest risk categories, while reducing screening for those in the lowest risk categories, potentially reducing overdiagnosis and costs while maintaining a reduced breast cancer death rate and improved life quality[7].

Many risk prediction algorithms have been developed to quantify the combined effect of various risk factors to predict the risk of developing breast cancer[8, 9]. The recently extended Breast and Ovarian analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) calculates cumulative risk of developing breast cancer based on family history, mammographic density, several lifestyle/hormonal and genetic risk factors[10]. BOADICEA includes the rare high to moderate risk pathogenic variants in breast cancer genes *BRCA1, BRCA2, PALB2, CHEK2* and *ATM,* and a Polygenic Risk Score (PRS) based on 313 breast cancer-associated variants (PRS_{313}). In 10 prospective studies, this PRS showed an association with breast cancer with an OR of 1.61 per standard deviation of the PRS distribution[11], and an area under receiver-operator curve of 0.630. It has been shown that the greatest breast cancer risk stratification in the general population and in women with a family history of breast cancer can be obtained by using the combined effects of the PRS and lifestyle/hormonal risk factors in the BOADICEA model[10].

Currently, breast cancer screening in the Dutch population is age-based[12]. Women start at age 50 years with biannual mammograms until the age of 75. Before considering risk-stratified approaches based on BOADICEA, it is important to assess its clinical validity in the Dutch population. In this study we validated the association between the PRS_{313} and breast cancer in a Dutch prospective cohort, its effect on predicting *in situ* breast cancer, and explore the discriminative ability of an individualised 10-year breast cancer risk score based on the PRS_{313} and several known risk factors using the BOADICEA version 5 model. We also assessed how a risk-based approach of population-based screening could have impacted breast cancer detection rates in our study cohort.

Materials and Methods

Study cohort

The Rotterdam Study (RS) is a prospective population-based cohort study of elderly Dutch individuals living in the Ommoord district of Rotterdam in the Netherlands[13]. Briefly, in the year 1989, individuals aged 55 or older were recruited into the RS-I cohort, which was extended in 2000 under similar criteria (RS-II-cohort) and in 2006 by the inclusion of individuals with an age between 45 and 55 (RS-III cohort). The overall response rate was 72%. In 2008 the Rotterdam Study comprised 14,926 subjects aged 45 years or older, including 8,823 women. For our study, we included all 6,670 women for whom genotype data were available. Genotyping was not performed for the excluded 2,153 women because of a low-quality DNA sample or because they declined blood-donation for DNA at study-entry.

Ethics statement

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus Medical Center and by the Dutch Ministry of Health, Welfare and Sports. All participants provided written informed consent to participate in the study and to have their medical information obtained from treating physicians.

Phenotype data

Diagnoses of cancer were collected for all individuals up to January 2014 and were based on medical records of general practitioners (including hospital discharge letters) and through linkage with Dutch Hospital Data, Netherlands Comprehensive Cancer Organisation, and histology and cytopathology registries in the region[13]. In total, 468 women had a breast cancer (invasive or *in situ*) diagnosis of whom 148 had been diagnosed prior to entry into the Rotterdam Study, and were excluded from further analyses. All participants were interviewed at home at inclusion, underwent extensive examinations every ~5 years in the Rotterdam Study research facility and received follow-up questionnaires (Figure S1), as described elsewhere[13]. Basic characteristics such as date of birth, vital status and age at inclusion were known for all participants. For most participants, information of breast cancer risk factors was available (Table S1, Total cohort), but family history of breast cancer and mammographic density were lacking. For the analyses, we used only information from the first questionnaire (Figure S1: RS-I-1, RS-II-1, RS-III-1) at the time of inclusion in the Rotterdam Study for variables that could vary over time, e.g. weight and alcohol use. Age at menopause was only included if menopause occurred before enrolment into the Rotterdam Study (Table S1, Subcohort).

Genotype data

Genotyping was performed with the Illumina 550K (RS-I and RS-II cohorts) and 610K (RS-III cohorts) arrays[13]. Standard quality control was completed, including selection on European ancestry, and imputation was performed using the Haplotype Reference Consortium (HRC) 1.1 and 1000G phase 3 reference panels[14, 15]. Of the 313 variants used to calculate the Polygenic Risk Score, 28 were directly genotyped by the arrays. Two variants were imputed with a quality below 0.3 and the remaining 283 variants were imputed with an average imputation quality of 0.95 (Table S2).

Polygenic Risk Score calculation

The following formula was used to calculate the PRS based on 313 variants:

$$PRS_j = \sum_{i=1}^{313} n_{ij} \ln (OR_i)$$

where n_{ij} is the number of risk alleles (0, 1 or 2) for variant i carried by individual j and OR_i is the per-allele odds ratio (OR) for breast cancer associated with variant i. The ORs were obtained from the Breast Cancer Association Consortium (BCAC) study[11] (Table S2). As the Estrogen Receptor (ER) status of the breast tumours was not available, only the overall breast cancer PRS was calculated. The PRS_{313} was standardised to the mean in all included women from the Rotterdam Study who did not develop incident breast cancer. To allow for direct comparison of PRS performance between both studies, the Standard Deviation (SD) of the population controls included in the validation-set from the BCAC study[11] was used, which was 0.609. For the calculations with BOADICEA version V, the PRS_{313} was standardised to the mean and SD from the population controls included in the total dataset from the BCAC study[11], which was -0.424 and 0.603 respectively.

Cumulative risk score calculation

Cumulative 10-year breast cancer risks were calculated with BOADICEA version V[10], starting at the age of inclusion in the Rotterdam Study, and using the birth-cohort incidence rates in combination with four different sets of variables, i.e., (i) age, (ii) age and PRS_{313}, (iii) age and risk factors, (iv) age, PRS_{313}, and risk factors. Risk factors included are age at menarche, age at menopause, number of children, age at first live birth, use of oral contraception, use of hormone replacement therapy, Body Mass Index (BMI), height, and alcohol use. For the variables that could vary over time, we used fixed variables. As BOADICEA ignores any risk factors for which the value is missing[10], no imputation was performed, and missing variables were kept missing.

Because BOADICEA calculates cumulative breast cancer risks up to age 80, 10-year breast cancer risks were only calculated for 4,377 women with an age of inclusion up to the age of 70 years. Women were considered affected if they developed breast cancer (invasive or *in situ*) within 10 years after inclusion in the Rotterdam Study.

Statistical analyses

Cumulative incidences were calculated using the Kaplan Meier method.

Association analyses

To estimate the association between the PRS_{313} and breast cancer risk in the Rotterdam Study cohort, Cox-regression analyses were performed. Relatedness among individuals of the same family was accounted for by correcting standard errors using a sandwich estimator. All models were adjusted by the age at inclusion in the Rotterdam Study. Incident breast cancer, *in situ* or invasive, was the event of interest. The time at risk was defined as the time elapsed between the inclusion date and the date of occurrence of the event of interest or right censoring. Right censoring could be due to (i) end of follow-up in January 2014 or (ii) death. The proportional hazard assumption for the model was tested. Sensitivity analyses were performed for (i) invasive breast cancer only by censoring the *in situ* breast cancer cases, (ii) *in situ* breast cancer only by censoring the invasive breast cancer cases, (iii) by censoring at the age of diagnosis of another type of cancer and (iv) by stratifying on Rotterdam Study cohort. To define the association between the PRS_{313} and other tumours than breast cancer, similar Cox-regression analysis was performed by censoring the breast cancer cases if they did not develop another tumour before the breast cancer diagnosis.

To investigate if the linearity assumption for the effect of PRS_{313} holds, we ran the model considering the categorical covariate given by the percentile groups of the PRS_{313} (0-10%; 10-20%; 20-40%; reference 40-60%; 60-80%; 80-90%; 90-100%) based on the distribution in the unaffected women in this cohort. The discrimination ability of the PRS_{313} in our sample was evaluated using the C-statistic[16], by groups based on quantiles of the age of inclusion in the Rotterdam Study (i.e. age <60, 60-70 and ≥70 years). Differences in the C-statistics were tested by computing bootstrap confidence intervals for the differences among groups.

Age-varying effect

The possible time-varying association of the PRS_{313} with breast cancer was investigated using age as time scale and considering three age dependent coefficients in the Cox model, corresponding to three different age intervals: (i) younger than 50 years, (ii) between 50 and 75 years old and (iii) above 75 years old. These cut-offs were chosen

based on their clinical relevance since women between 50 and 75 years are eligible for population screening according to the Dutch guideline[12].

Clinical validity of BOADICEA v5

To validate the BOADICEA 10-year cumulative risk scores, model calibration and discrimination ability in our sample were assessed. Calibration was investigated by comparing overall observed versus expected cumulative risks and by visually inspecting the calibration plots based on risk deciles. Because of the presence of right censoring, empirical risks at 10 years were estimated using the Kaplan-Meier method. As in the association analyses, discrimination was evaluated using C-statistics.

Statistical significance was defined as a two-sided p-value of <0.05. All analyses were performed with R version 3.5.3.[17]

Results

We included 6,522 women in the main analyses with an average age at study-entry of 66 years. Of these, 320 developed either invasive or *in situ* breast cancer during follow-up and 744 developed another type of tumour; the overlap between these two groups was 16, all of whom developed another type of tumour first (Table S3). The median follow-up calculated with the reverse Kaplan-Meier method was 12.40 years, with a minimum and maximum follow up of 0.03 and 24.43 years. Cohort characteristics are shown in Table S1. The average PRS_{313} in groups of affected (i.e. invasive, *in situ*, and a second breast tumour) and unaffected women (including women who developed another tumour than breast cancer) are shown in Figure S2 and Table S4.

Breast cancer cumulative incidence

The cumulative incidence of breast cancer in the total cohort was on average 4.2%, 95%CI [3.7%-4.8%] and 7.3%, 95%CI [6.4%-8.2%] 10 and 20 years after inclusion respectively. Stratified by quintiles of the PRS_{313}, after 20 years of follow-up, the incidence in the highest quintile was 10.8%, 95%CI [8.5%-13.1%] and 4.4%, 95%CI [2.8%-6.0%] for the lowest quintile (Figure S3).

Association analyses

A significant association was found between the PRS_{313} and incident breast cancer with an HR per SD of 1.56, 95%CI [1.40-1.74], $p=2.47x10^{-15}$ (Table 1). There was no evidence of violation of the proportional hazard assumption (p-value=0.716), indicating that the HR remained constant over time. The discriminative ability of the PRS_{313}, as measured by the C-statistic, was 0.632, 95%CI [0.58-0.69], 0.673, 95%CI [0.61-0.73], and 0.562, 95%CI

[0.48-0.62] for women included before age 60, between age 60 and 70, and above age 70 respectively (Table 1).

Sensitivity analyses for (i) invasive breast cancer only, (ii) censoring at another tumour if applicable or (iii) stratifying by the Rotterdam Study subcohort all showed similar results (Table 1). Notably, also i*n situ* breast cancer showed a statistically significant association with the PRS_{313}, HR per SD=1.43, 95%CI [1.01-2.01], p=0.042.

Association analyses for breast cancer and percentiles of the PRS_{313} showed that the HR-estimates were in line with the HR predicted when a continuous PRS_{313} is assumed, under a log-linear model (Figure 1, Table 1).

During follow-up, 744 women developed another tumour than breast cancer without evidence for association with the PRS_{313} (HR per SD=1.05, 95%CI [0.98-1.12], p-value=0.195).

Age-varying effect

Extension of the Cox model allowing for age-dependent regression coefficients showed that the performance of the PRS_{313} decreased with increasing inclusion age, with the HRs per SD declining from 2.74, 95%CI [1.72-4.37] for women included before age 50, to 1.74, 95%CI [1.52-2.00] for women included between 50 and 75 (p_{diff}=0.066). The HR for women included after age 75 was 1.29, 95%CI [1.08-1.55], and the p-value of the difference with respect to the youngest group was 0.003 (Table 1).

Table 1: Results of the association analyses between breast cancer and the PRS_{313}

		n Included	n Events	HR	95% CI	p-value	C-statistic[c]	95% CI
Main analyses		6522	320	1.56	1.40-1.74	$2.47x10^{-15}$		
Age category for discriminative ability of the PRS	**<60**	2175	104				0.632	0.58-0.69
	60-70	2174	128				0.673	0.61-0.73
	≥70	2173	88				0.562	0.48-0.62
Sensitivity analyses	**Invasive BC only**	6522	290[a]	1.57	1.40-1.77	$1.34x10^{-14}$		
	In situ BC only	6522	34	1.43	1.01-2.01	0.042		
	Censored at other tumour	6402[b]	298	1.54	1.37-1.73	$1.88x10^{-13}$		
	Stratified by RS cohort	6522	320	1.56	1.40-1.75	$1.92x10^{-15}$		
Percentage of the PRS	**0-10%**	637	17	0.59	0.34-1.01	0.053		
	10-20%	636	16	0.58	0.33-1.01	0.053		
	20-40%	1283	42	0.73	0.49-1.09	0.120		
	40-60%	1298	57	1.00	ref	ref		
	60-80%	1325	85	1.49	1.07-2.09	0.019		
	80-90%	656	36	1.28	0.84-1.94	0.251		
	90-100%	687	67	2.37	1.66-3.37	$1.73x10^{-06}$		
Age category for time-varying analyses	**<50**	224	2	2.74	1.72-4.37	$2.23x10^{-05}$		
	50-75	5104	197	1.74	1.52-2.00	$2.21x10^{-15}$		
	>75	4032	121	1.29	1.08-1.55	0.005		

Abbreviations: BC, Breast Cancer; CI, Confidence Interval; HR, Hazard Ratio; PRS, Polygenic Risk Score; RS, Rotterdam Study.

[a] 4 women developed an invasive breast tumour after development of an *in situ* breast tumour.

[b] 120 women were excluded from analyses because they developed another tumour before inclusion in the Rotterdam study.

[c] The corresponding differences in C-statistic were for women with inclusion age 60-70 versus age <60: 0.041, 95%CI [-0.05-0.12]; for women with inclusion age 60-70 versus age ≥70: 0.111, 95%CI [0.02-0.21]; for the women with inclusion age <60 versus age ≥70: 0.070, 95%CI [-0.01-0.18].

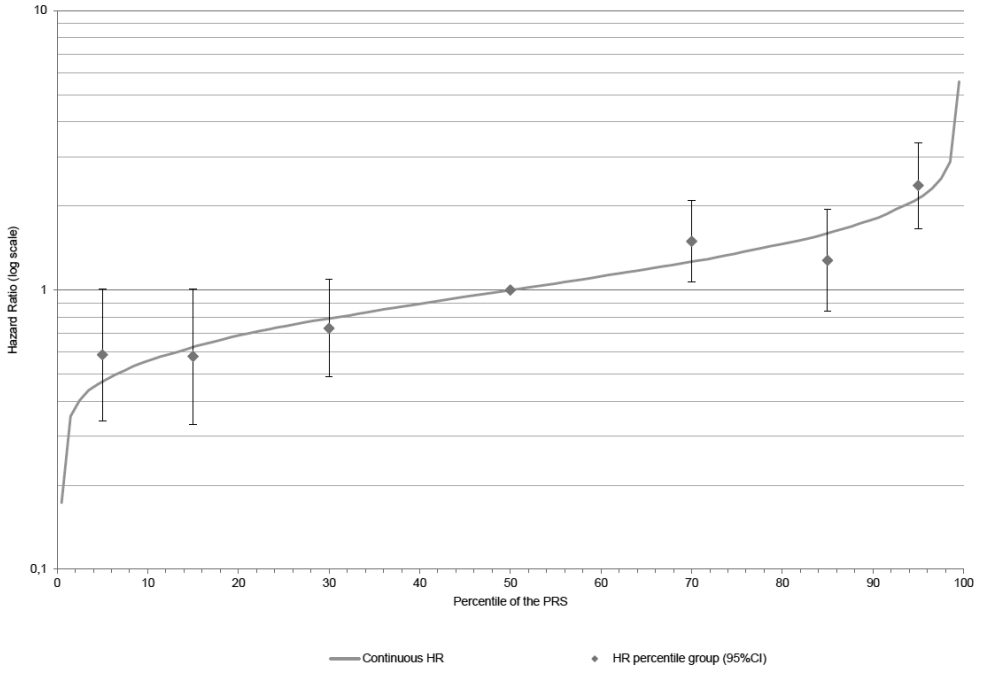


Figure 1: Association with the PRS_{313} and breast cancer risk
Plot of the HR for the association between the PRS_{313} and breast cancer risk based on PRS_{313} percentiles. The PRS_{313} percentile groups are 0-10%, 10-20%, 20-40%, 40-60% (reference), 60-80%, 80-90%, 90-100% based on the distribution in unaffected women. The numbers and corresponding effect sizes are shown in Table 1. The solid line represents the continuous distribution based on the per SD effect size of the PRS_{313}.
Abbreviations: CI, Confidence Interval; HR, Hazard Ratio; PRS, Polygenic Risk Score.

Clinical validity of BOADICEA V5

For these analyses, we selected 4,377 women with an age of inclusion under 70 years. Of these, 163 developed breast cancer within 10 years after inclusion, of whom 142 invasive. The median follow-up in this subcohort was 10 years (range 0.03 – 10 years), and the cumulative incidence of breast cancer was 4.4% (95%CI [3.7%-5.1%]). The distributions of 10-year cumulative risk scores under different models are shown in Figure S4. Irrespective of the variables included, BOADICEA underestimated the observed risk of 4.4% (Table 2). Accordingly, while using age and PRS_{313} seems to result in the best calibration (Figure S4C), it underestimated the observed risks in the higher risk categories. The highest discriminative ability was found for the model with age, PRS_{313} and all available risk factors (0.653, 95%CI [0.60-0.70]), henceforth the "full" model. The PRS_{313} was the strongest factor contributing to discrimination, relative to age and other risk factors (Table 2).

Using the full model and a threshold of 2.5% 10-year breast cancer risk, which approximates the risk of women entering the age-based population screening program in The Netherlands, 101 cases (62% of incident cases) occurred in a screening-group of 1,956 women (45% of total) and 2,421 women would not be screened, in which 62 breast cancers occurred (Figure 2; Table 3). Using the PRS_{313} and age only, 130 cases (80% of incident cases) occurred in a screening-group of 2,863 women (65% of total); 1,481 women would not be screened, in which 33 breast cancers occurred. In Figure S6 the percentages of incident breast cancer cases and unaffected women are shown for different category thresholds. For both models, the invasive cancers in the group selected for screening were more likely to be of lower grade compared to the cancers in the non-screened group (Table 3). The reverse effect was found for *in situ* cancers.

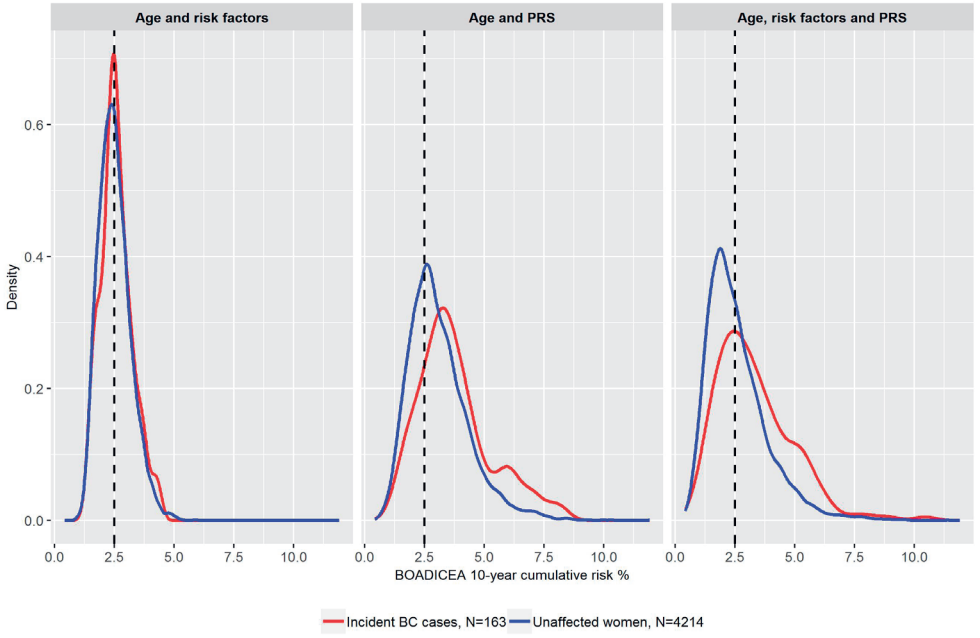


Figure 2: Cumulative 10-year breast cancer risk distribution predicted by BOADICEA
Density plots of the cumulative 10-year risk calculated by BOADICEA for unaffected women and incident breast cancer cases. Including age and risk factors (left), including age and the PRS_{313} (middle) and the full model including age, risk factors and the PRS_{313}. The dashed line shows the threshold of a 10-year risk of 2.5%.
Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PRS, Polygenic Risk Score.

Table 2: Range and discriminative ability of the cumulative 10-year breast cancer risk scores calculated with BOADICEA

Variables included	Mean % (range)		C-statistic	95%CI
	Unaffected women	BC cases[a]		
Age	3.0 (2.2-3.6)	2.9 (2.2-3.6)	0.531	0.50-0.58
Age, risk factors	2.5 (1.0-5.9)	2.6 (1.4-4.3)	0.558	0.52-0.60
Age, PRS_{313}	3.1 (0.6-11.9)	3.8 (1.2-8.3)	0.636	0.59-0.68
Age, risk factors, PRS_{313}	2.6 (0.4-11.4)	3.3 (0.9-10.5)	0.653	0.60-0.70

Abbreviations: BC, Breast Cancer; CI, Confidence Interval; PRS, Polygenic Risk Score
[a] Women who developed BC within 10 years of follow up.

Table 3: Numbers and percentages of women per 10-year risk category

			Total	10-year risk category based on BOADICEA			
				Including age and PRS		Including age, risk factors and PRS	
				<2.5%	>2.5%	<2.5%	>2.5%
Unaffected women	All		4214	1481 (35%)	2733 (65%)	2359 (56%)	1855 (44%)
Incident BC cases	All		163	33 (20%)	130 (80%)	62 (38%)	101 (62%)
	invasive BC	All	142	30 (21%)	112 (79%)	52 (37%)	90 (63%)
		Grade 1	19	2 (11%)	17 (89%)	3 (16%)	16 (84%)
		Grade 2	38	7 (18%)	31 (82%)	12 (32%)	26 (68%)
		Grade 3	43	13 (30%)	30 (70%)	21 (49%)	22 (51%)
		Unknown	42	8 (19%)	34 (81%)	16 (38%)	26 (62%)
	In situ BC	All	21	3 (14%)	18 (86%)	10 (48%)	11 (52%)
		Grade 1	3	2 (67%)	1 (33%)	2 (67%)	1 (33%)
		Grade 2	3	1 (33%)	2 (67%)	2 (67%)	1 (33%)
		Grade 3	13	0 (0%)	13 (100%)	5 (38%)	8 (62%)
		Unknown	2	0 (0%)	2 (100%)	1 (50%)	1 (50%)

Abbreviations: BC, Breast Cancer; BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PRS, Polygenic Risk Score

Discussion

Many risk factors for breast cancer, both genetic and non-genetic, have been identified the past decades[18, 19]. Increasingly, these are being integrated into computational models that allow personalised breast cancer risk assessment, which has potential application beyond current practice of genetic testing in family cancer clinics[8, 9, 20]. The BOADICEA algorithm is among the most comprehensive risk models presently available for breast cancer risk assessment[10]. Here, we validated the most recent version of this model in a large prospective population-based Dutch cohort of women above 45 years, which hasn't been part of the previously published BCAC study[11]. Unsurprisingly, the best discrimination was achieved after inclusion of all available risk factors, with the largest contribution deriving from the PRS_{313}. The PRS_{313} was significantly associated with breast cancer, with a similar effect size as in other prospective series of different geographic origin[11], demonstrating its robustness and potential application to the Dutch population.

The PRS_{313} improved the discriminatory ability from 0.531 to 0.636, compared with a model using age only, which could only be marginally improved further (to 0.653) by adding lifestyle, reproductive factors and anthropometric data. This is in line with previous research, showing that the variance explained by the risk factors are modest compared to the PRS_{313} risk stratification[10, 21]. Results of the calibration showed that BOADICEA underestimated the observed risks, especially in the higher categories of risk. One possible explanation is that BOADICEA v5 uses the population breast cancer incidences of the United Kingdom as baseline risk, which are slightly lower than those in the Netherlands[1]. But more importantly, data on family history, mammographic density and rare high-risk variants in *BRCA1* and *BRCA2* were lacking in our cohort. In another prospective validation study of a previous version of BOADICEA in two cohorts of women from Australia, Canada, and the USA, information on family history and *BRCA1/2* carrier status, but not the PRS_{313}, was available, and here, BOADICEA overestimated 10-year cumulative risks in the highest risk quantile[9]. Possibly, the missing data on family history and *BRCA1/2* status in the Rotterdam Study were in fact more prevalent than modelled by BOADICEA. Our calibration results indicate that for proper use in the general population, information on family history may be important.

We illustrated the potential impact of the model in detecting breast cancer in a population-screening setting in which women would participate based on their individual risk. In this illustration, the PRS_{313} alone would have detected more cases than the full BOADICEA model, but would also have identified a larger screening group. Apparently, women in the Rotterdam Study have on average fewer non-genetic risk factors compared to the total population, which on average slightly modifies their risk in a downward direction. The PROCAS study used the Tyrer-Cuzick model with mammographic density and risk factors,

combined with a PRS based on 18 SNPs[22]; they found 82% of the cases to occur in 68% of women with a 10-year breast cancer risk above 2%, i.e., very similar to what we found with the PRS_{313} alone.

Remarkably, we found the proportion of low grade invasive tumours to be higher in those with a 10-year risk >2.5%, compared to those with lower risks. Screen-detected invasive cancers are more likely of lower grade and stage[23]. Our cohort data did not include information on whether incident breast cancers were screen-detected or not, hence we cannot exclude that high-risk women disproportionally self-selected for mammographic screening, which could explain this bias. In contrast, for the *in situ* carcinomas, more high grade tumours were found in the >2.5% 10-year risk group compared to those with lower risks. Histological grade of Ductal Carcinoma In Situ (DCIS) has been suggested to be one of six factors associated with subsequent development of invasive disease[24], albeit not very strongly so. It remains possible that the PRS_{313} is more strongly associated with low grade invasive breast cancer than with higher grades, as observed for some individual variants[25, 26], and inversely so for DCIS. It will be important to replicate this in larger studies to inform the evaluation of the cost-effectiveness of a risk-based versus age-based entry of the population-screening[7].

Although PRS development studies have included only invasive breast cancer[11, 27], in our cohort the PRS_{313} is associated with i*n situ* breast cancer as well, with a non-significantly lower effect-size than for invasive breast cancer. This corresponds well with a previously reported association of an 18-SNP-based PRS[22] and with previous results showing that the association of 51 of the 76 investigated breast cancer loci with DCIS is in the same direction as for invasive breast cancer[28]. Although BOADICEA is presented as a model that predicts invasive breast cancer[10], these results suggest it might also predict *in situ* breast cancer. Larger studies are needed to confirm this and provide more accurate risk estimates, specifically in the setting of population screening programs.

As in previous studies[11, 27], we found that the effect-size of the PRS for breast cancer declined with increasing age. While this is not yet modelled in BOADICEA, this could be important to consider for women under the age of 50 who are at this moment not eligible for population breast cancer screening in the Netherlands, because our results suggest that using the overall HR would be underestimating risk in this age group.

In the Rotterdam Study, malignancies other than breast cancer are also recorded. We found no evidence for association of the PRS_{313} with these cancers, suggesting it specifically predicts breast cancer. Another prospective study also reported no association between other types of cancer and a sum of breast cancer risk alleles at 72 loci[29]. Because we only

analysed all other tumours combined, we cannot exclude that the PRS_{313} has an association with one specific type of other cancer.

A strength of our study is the prospective population-based study design, including all women in a specified locale near Rotterdam. Because of the high response rate (>70%) it is a good representation of the Dutch population in that age category[13]. Furthermore, for a large group of women, there is extensive follow up of up to 25 years.

Besides that information on mammographic density and family history was lacking, another limitation of our study is the unknown ER-status of the breast tumours, precluding the analysis of ER-positive and ER-negative disease separately. Furthermore, to evaluate the introduction of risk-based entry into population-screening, establishing the detection rate of breast cancers below the age of 50 would have been relevant, which was not possible in our older cohort of women. Finally, we excluded nearly 25% of all women in the Rotterdam Study because no genotyping data were available. Declining blood-donation for DNA extraction did not lead to differences in the basic characteristics between the genotyped and non-genotyped groups. Therefore, if a selection bias was present, we believe this bias would be small.

In summary, the PRS_{313} replicates robustly in the Dutch population and the discriminative power of the BOADICEA model seems appropriate for implementation into breast cancer prevention programs, such as those currently ongoing in cancer family clinics in many countries worldwide. However, application to the general population would require recalibration of BOADICEA to address underestimation in the higher risk categories. Although the Rotterdam Study design precluded analysis of breast cancer-specific mortality, our evaluation of clinical validity provides first insights into how a risk-based entry could impact the efficacy of the breast cancer population screening program in the Netherlands.

Acknowledgment and funding

This work was supported by the Dutch Cancer Society (KWF), grant UL2014-7473.

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants,

the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

The generation and management of GWAS genotype data for the Rotterdam Study (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data.

Disclosure of potential conflicts of interest

The authors declare no conflicts of interest

References

1. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European journal of cancer (Oxford, England : 1990)*. Nov 2018;103:356-387. doi:10.1016/j.ejca.2018.07.005
2. van der Waal D, Verbeek AL, den Heeten GJ, Ripping TM, Tjan-Heijnen VC, Broeders MJ. Breast cancer diagnosis and death in the Netherlands: a changing burden. *European journal of public health*. Apr 2015;25(2):320-4. doi:10.1093/eurpub/cku088
3. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *British journal of cancer*. Jun 11 2013;108(11):2205-40. doi:10.1038/bjc.2013.177
4. Ripping TM, Verbeek AL, Fracheboud J, de Koning HJ, van Ravesteyn NT, Broeders MJ. Overdiagnosis by mammographic screening for breast cancer studied in birth cohorts in The Netherlands. *International journal of cancer*. Aug 15 2015;137(4):921-9. doi:10.1002/ijc.29452
5. The benefits and harms of breast cancer screening: an independent review. *Lancet (London, England)*. Nov 17 2012;380(9855):1778-86. doi:10.1016/s0140-6736(12)61611-0
6. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Annals of internal medicine*. Oct 18 2011;155(8):481-92. doi:10.7326/0003-4819-155-8-201110180-00004
7. Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. *JAMA oncology*. Nov 1 2018;4(11):1504-1510. doi:10.1001/jamaoncol.2018.1901
8. 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
9. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. *The Lancet Oncology*. Apr 2019;20(4):504-517. doi:10.1016/s1470-2045(18)30902-1
10. 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
11. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
12. IKNL. Netherlands Comprehensive Cancer Organisation: Oncoline Mammacarcinoom. 2019. Available from www.oncoline.nl/richtlijn/item/index.php?pagina=/richtlijn/item/pagina.php&richtlijn_id=885. Accessed December, 2019;
13. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *European journal of epidemiology*. May 4 2020;doi:10.1007/s10654-020-00640-5

14. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*. Oct 2016;48(10):1279-83. doi:10.1038/ng.3643
15. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. Oct 1 2015;526(7571):68-74. doi:10.1038/nature15393
16. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*. May 10 2011;30(10):1105-17. doi:10.1002/sim.4154
17. R_Core_Team_(2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
18. Rojas K, Stuckey A. Breast Cancer Epidemiology and Risk Factors. *Clinical obstetrics and gynecology*. Dec 2016;59(4):651-672. doi:10.1097/grf.0000000000000239
19. Lakeman IMM, Schmidt MK, van Asperen CJ, Devilee P. Breast Cancer Susceptibility—Towards Individualised Risk Prediction. journal article. *Current Genetic Medicine Reports*. June 01 2019;7(2):124-135. doi:10.1007/s40142-019-00168-5
20. Turnbull C, Sud A, Houlston RS. Cancer genetics, precision prevention and a call to action. *Nature genetics*. Sep 2018;50(9):1212-1218. doi:10.1038/s41588-018-0202-0
21. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA oncology*. Oct 1 2016;2(10):1295-1302. doi:10.1001/jamaoncol.2016.1025
22. 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
23. Autier P, Boniol M, Koechlin A, Pizot C, Boniol M. Effectiveness of and overdiagnosis from mammography screening in the Netherlands: population based study. *BMJ (Clinical research ed)*. Dec 5 2017;359:j5224. doi:10.1136/bmj.j5224
24. Visser LL, Groen EJ, van Leeuwen FE, Lips EH, Schmidt MK, Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. May 2019;28(5):835-845. doi:10.1158/1055-9965.Epi-18-0976
25. Garcia-Closas M, Hall P, Nevanlinna H, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS genetics*. 4/25/2008 2008;4(4):e1000054. In File.
26. Purrington KS, Slettedahl S, Bolla MK, et al. Genetic variation in mitotic regulatory pathway genes is associated with breast tumor grade. *Hum Mol Genet*. 11/15/2014 2014;23(22):6034-6046. In File.
27. 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]

28. Petridis C, Brook MN, Shah V, et al. Genetic predisposition to ductal carcinoma in situ of the breast. *Breast cancer research : BCR*. Feb 17 2016;18(1):22. doi:10.1186/s13058-016-0675-7
29. Naslund-Koch C, Nordestgaard BG, Bojesen SE. Common breast cancer risk alleles and risk assessment: A study on 35,441 individuals from the Danish general population. *Annals of oncology : official journal of the European Society for Medical Oncology*. Oct 13 2016;doi:10.1093/annonc/mdw536

Supplementary figures and tables

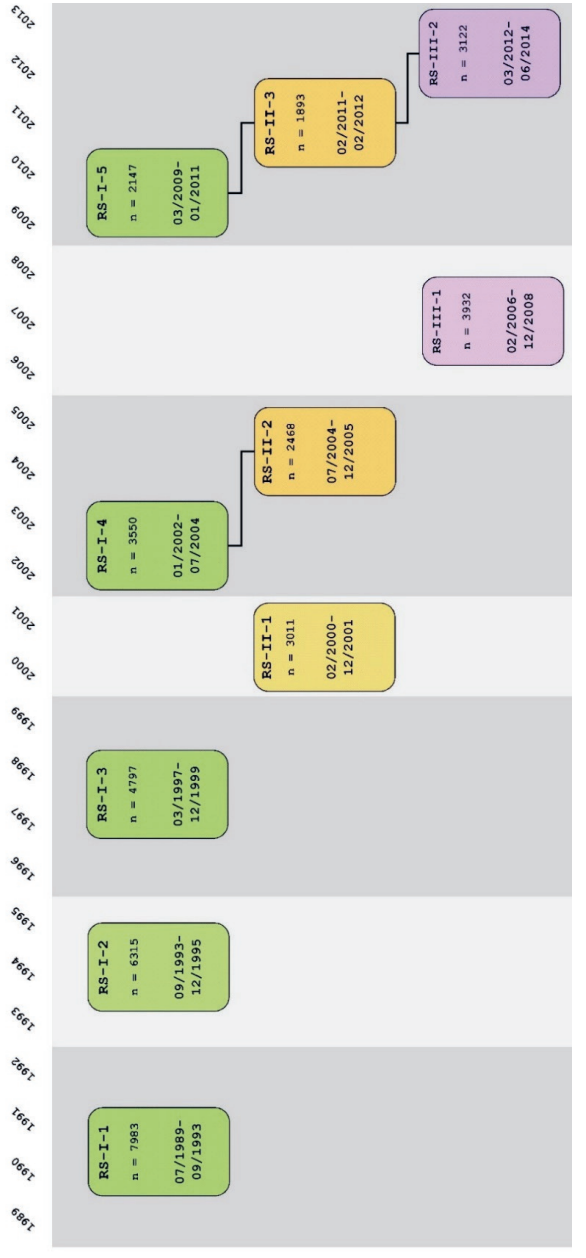


Figure S1: Time points of the first and follow up questionnaires for the Rotterdam Study cohorts
Diagram of the included individuals and contact moments for the Rotterdam Study. Green boxes: RS-I cohort, yellow boxes: RS-II cohort; purple boxes: RS-III cohort. A more elaborate figure with information about the follow up after 2013 is published by Ikram et al.[1].

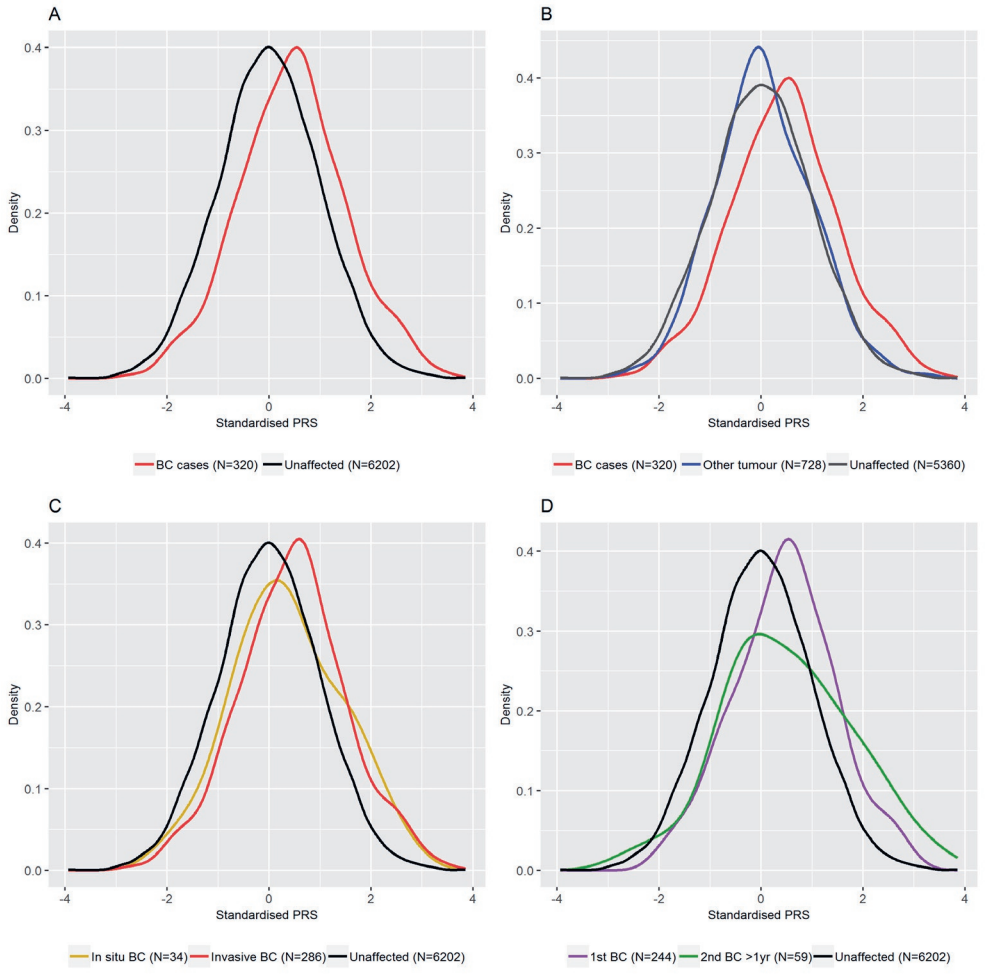


Figure S2: Distribution curves of the PRS_{313} in the Rotterdam Study cohort

Abbreviations: BC, Breast Cancer; PRS, Polygenic Risk Score

The standardised PRS_{313} was plotted against the density for different groups in the Rotterdam Study. (A) incident BC cases and unaffected women; (B) incident BC cases, unaffected women who developed another type of tumour and unaffected women who did not develop another type of tumour. Women who developed another type of tumour before inclusion in the Rotterdam Study were excluded (N=114); (C) invasive incident BC cases, *in situ* incident BC cases and unaffected women; (D) Incident BC cases who developed one breast tumour, incident BC cases who developed a second primary breast tumour after one year and unaffected women. Women who developed a second primary breast tumour within one year were excluded (N=17). Unaffected women include all those that did not develop BC.

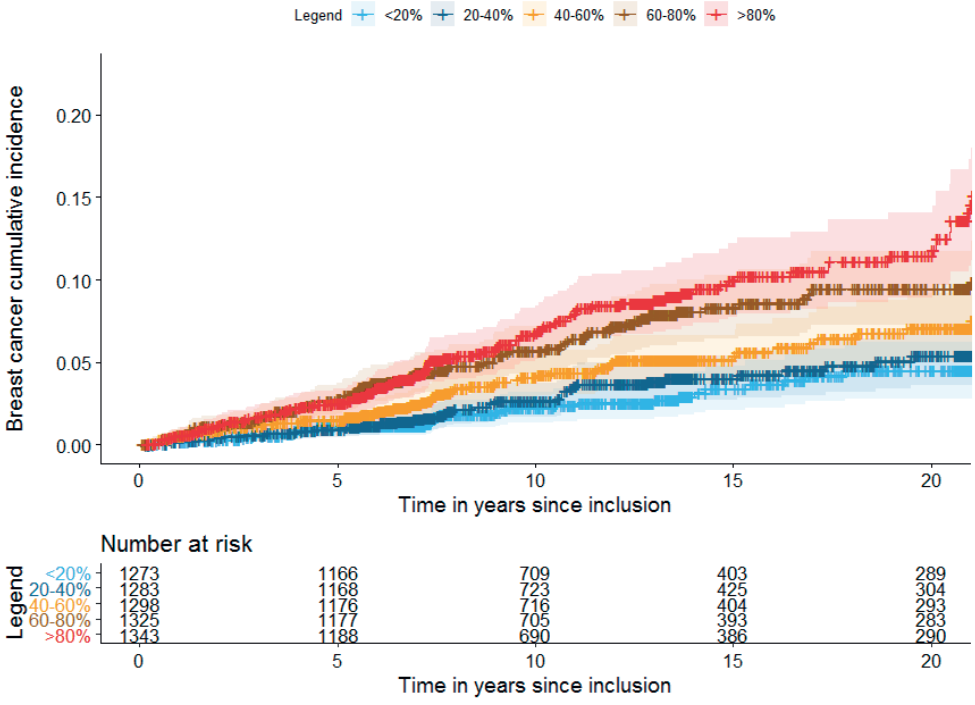


Figure S3: Cumulative breast cancer incidence in the Rotterdam study stratified on PRS_{313} quintiles

Abbreviations: PRS, Polygenic Risk Score.

Kaplan Meier plot for the cumulative breast cancer incidence since the time of inclusion in the Rotterdam Study. The cohort is stratified in quintiles of the PRS_{313}, based on the distribution of unaffected women in the cohort.

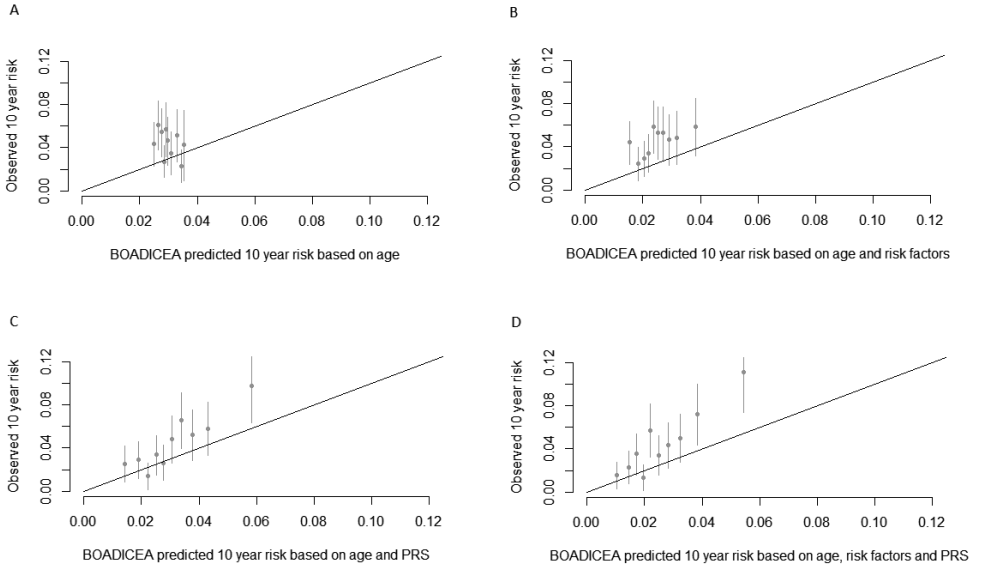


Figure S4: Calibration plots of the predicted 10-year risk based on BOADICEA and the observed risk in the Rotterdam Study cohort

Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PRS, Polygenic Risk Score.

10-year cumulative BC risks were calculated for all women included in the Rotterdam Study before the age of 70 years, using BOADICEA v5. The difference between the observed and predicted risk is shown per decile of the predicted risk, including 95% confidence intervals, for different sets of included variables. Using age only (A), age and risk factors (B), age and the PRS (C) and age, risk factors, and the PRS (D).

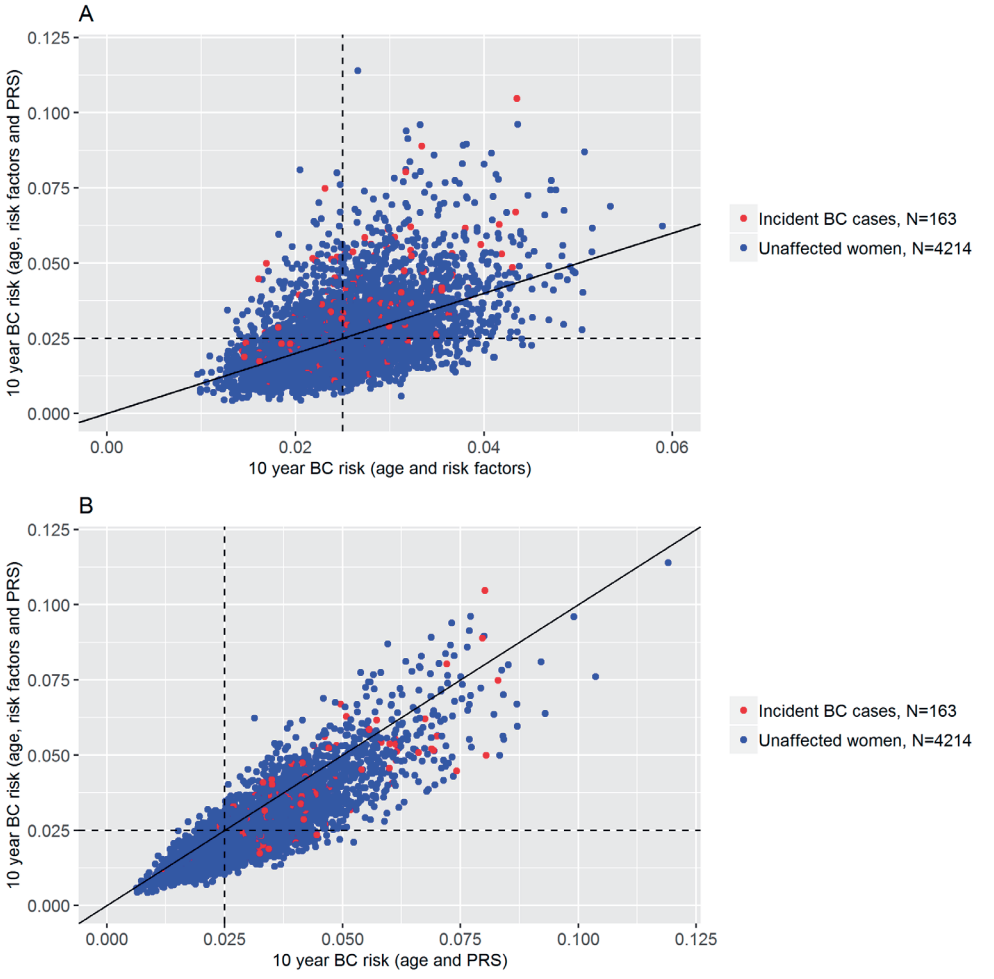


Figure S5: Change in 10-year risk by adding risk factors or the PRS_{313} in the BOADICEA model
Abbreviations: BC, Breast Cancer; BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PRS, Polygenic Risk Score.
10-year cumulative BC risks were calculated for all women included in the Rotterdam Study before the age of 70 years, using BOADICEA v5. Women were considered as incident BC cases if they developed BC within 10 years of follow up (shown in red). (A) Risk-change by adding the PRS_{313} in the BOADICEA model (y-axis) including age and risk factors (x-axis). (B) Risk-change by adding risk factors in the BOADICEA model (y-axis) including age and the PRS_{313} (x-axis).

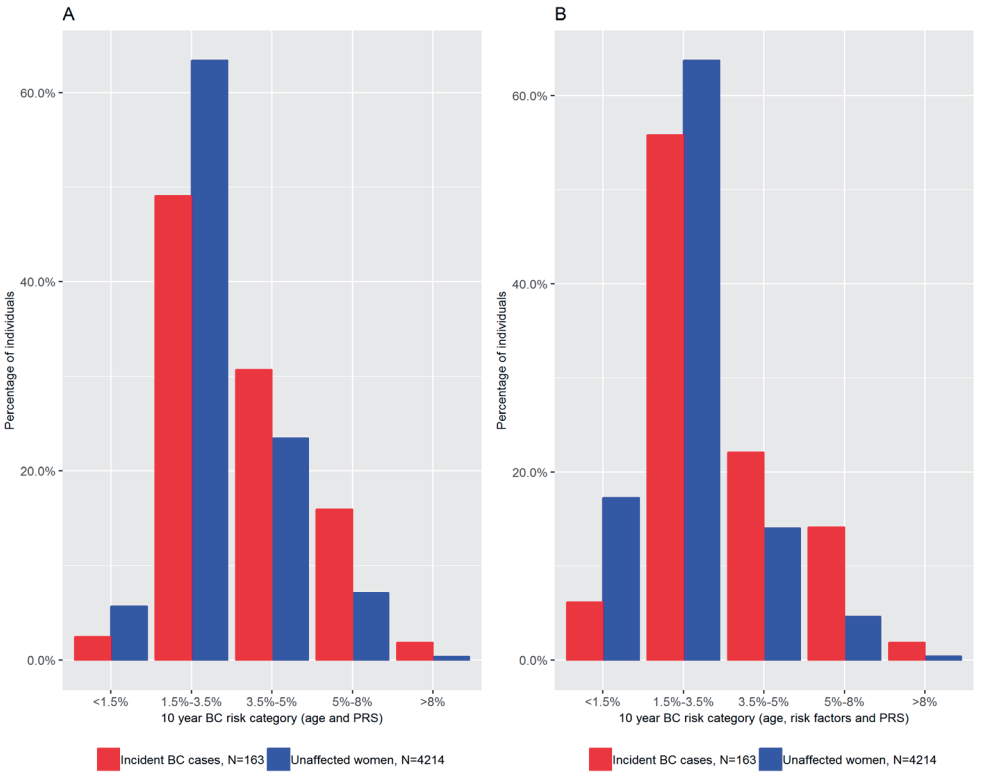


Figure S6: Percentage of unaffected women and incident breast cancer cases in different 10-year risk categories

Abbreviations: BC, Breast Cancer; BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PRS, Polygenic Risk Score.

Bar plot of the percentages of women assigned to the different 10-year cumulative BC risk categories (<1.5%; 1.5%-3.5%; 3.5%-5%; 5%-8%; >8%) as calculated with BOADICEA v5 using two sets of variables. Including age and the PRS_{313} (A) and including age, risk factors and the PRS_{313} (B). These risks were calculated for all women included in the Rotterdam Study below the age of 70 years. Women were considered affected if they developed BC within 10 years of follow up.

Table S1: Characteristics of the Rotterdam Study cohort

		Total cohort		Subcohort[a]	
		Unaffected	Incident BC	Unaffected	Incident BC
Number		6202	320	4214	163
Rotterdam Study cohort	**RS-I**	3536	227	1821	152
	RS-II	1057	59	796	50
	RS-III	1609	34	1525	33
Birth cohort	**<1900**	54	1	0	0
	1900-1910	487	9	0	0
	1910-1920	996	46	0	0
	1920-1930	1441	106	976	77
	1930-1940	1293	97	1235	97
	1940-1950	1087	42	1087	82
	1950-1960	811	18	811	18
	1960	33	1	33	1
Age at inclusion	**Mean**	66.1	65	59.9	60.4
	Range	45.8-99.2	45.8-96.3	45.8-70.0	45.8-70.0
Age at diagnosis	**Mean**	-	72.7	-	65.3
	Range	-	48-100	-	48.0-79.0
Invasiveness first BC	**Invasive**	-	286	-	142
	In situ	-	34	-	21
Asynchronous second BC[b]	**All**	-	59	-	44
	Invasive	-	59	-	44
	In situ	-	0	-	0
Other incident tumour[c]		728	16	450	13
Risk factors					
Height in cm	**Mean**	162.3	163.0	164.0	164.3
	Unknown	137 (2%)	5 (2%)	9 (0.2%)	3 (2%)
Alcohol use in grams per day	**Mean**	6.3	7.1	6.8	6.8
	Unknown	506 (8%)	11 (3%)	742 (18%)	34 (21%)
Age menarche	**Mean**	13.5	13.3	13.4	13.3
	Unknown	317 (5%)	11 (3%)	102 (2%)	4 (2%)
Age menopause	**Mean**	48.8	49.2	48.6	49.4
	Unknown	473 (8%)	24 (8%)	255 (6%)	15 (9%)
	Premenopausal	-	-	187	5
Number of children	**0**	482	25	408	15
	1	811	39	642	22
	2	1819	93	1549	52
	>2	1443	80	1031	41
	Unknown	1647 (27%)	83 (26%)	584 (14%)	33 (20%)
Age at first childbirth	**Mean**	25.2	25.2	25.0	25.6
	Unknown[d]	47 (1%)	5 (2%)	603 (14%)	34 (21%)

Use of oral contraception	**Never**	2346	137	1238	50
	Ever	2774	126	2665	90
	Unknown	1082 (17%)	57 (18%)	311 (7%)	23 (14%)
Use of hormone replacement therapy	**Never**	5050	254	3416	128
	Ever	994	62	758	32
	Unknown	158 (2.5%)	4 (1%)	40 (1%)	3 (2%)
Body Mass Index	**Mean**	27.0	27.7	27.0	27.8
	Unknown	141 (2%)	5 (2%)	38 (1%)	3 (2%)
Standardised PRS_{313}	**Mean**	0	0.45	-0.01	0.57
	SD	1.00	1.05	1.00	1.02

Abbreviations: BC, Breast Cancer; PRS, Polygenic Risk Score; RS, Rotterdam Study; SD, Standard Deviation.

[a] Subcohort of women with an age of inclusion in the Rotterdam Study up to age 70

[b] Development of a second primary breast tumour at least one year after the first primary breast tumour.

[c] For women who developed BC during follow up, other tumours were only reported in this study if the other tumour was diagnosed before the BC diagnosis.

[d] For women known to have children.

Table S2: 313 breast cancer associated variants included in the Polygenic Risk Score

See online material. First 7 columns of the table are published by Mavaddat et al.[2]

Table S3: Number of included women diagnosed with other type of tumours

ICD10	Tumour description[a]	Unaffected women	Incident BC cases[b]	Total
C00	Lip	5		5
C02	Tongue	2		2
C03	Gum	1		1
C04	Floor of mouth	1		1
C05	Palate	2		2
C06	Mouth	2		2
C08	Major salivary glands	1		1
C09	Tonsil	2		2
C10	Oropharynx	1		1
C15	Oesophagus	27		27
C16	Stomach	21		21
C17	Small intestine	3	1	4
C18	Colon	90	1	91
C19	Rectosigmoid	33	2	35
C20	Rectum	38	1	39
C21	Anus and anal canal	5		5
C22	Liver and intrahepatic bile ducts	8		8
C23	Gallbladder	2		2
C24	Biliary tract	6		6
C25	Pancreas	44		44
C26	Digestive organs	4		4
C32	Larynx	1		1
C34	Bronchus & lung	112		112
C39	Respiratory system and intrathoracic organs	1		1
C40	Bone and articular cartilage of limbs	2		2
C43	Melanoma	27	2	29
C45	Mesothelioma	4		4
C48	Retroperitoneum and peritoneum	1		1
C49	Connective and soft tissue	3		3
C51	Vulva	6		6
C52	Vagina	1		1
C53	Cervix uteri	10	1	11
C54	Corpus uteri	48	2	50
C56	Ovary	24		24
C57	Female genital organs	1		1
C64	Kidney, except renal pelvis	16	1	17
C65	Renal pelvis	4		4
C66	Ureter	1		1
C67	Bladder	24	1	25
C69	Eye and adnexa	5		5
C70	Meninges	1		1
C71	Brain	13		13
C73	Thyroid gland	4	1	5
C80	Malignant neoplasm unspecified	37	1	38

C81	Hodgkin lymphoma	1		1
C82	Follicular lymphoma	6	1	7
C83	Non-follicular lymphoma	13		13
C84	Mature T/NK-cell lymphomas	2		2
C85	Non-Hodgkin lymphoma	11		11
C88	Immunoproliferative diseases	1		1
C90	Multiple myeloma and malignant plasma cell neoplasms	19	1	20
C91	Lymphoid leukaemia	12		12
C92	Myeloid leukaemia	17		17
C93	Monocytic leukaemia	2		2
Total		728	16	744

Abbreviations: BC, Breast Cancer; ICD, International Classification of Diseases and Related Health Problems

[a] ICD10 tumour description[3]

[b] Other tumours are only reported if a woman developed this tumour before the BC diagnosis

Table S4: Descriptives for the standardised PRS_{313}

		Number	Mean	SD	SE	95% CI
Unaffected	**Total**	6202	0.00	1.00	0.01	-0.02-0.02
	Without other tumour	5360	-0.01	1.01	0.01	-0.03-0.02
	Incident other tumour[a]	728	0.03	0.98	0.04	-0.04-0.10
Incident BC cases	**Total**	320	0.45	1.05	0.06	0.34-0.57
	Invasive BC	286	0.46	1.05	0.06	0.34-0.58
	***In situ* BC**	34	0.36	1.06	0.18	0.00-0.72
	One primary breast tumour	244	0.46	1.00	0.06	0.33-0.59
	Asynchronous second BC[b]	59	0.51	1.27	0.17	0.19-0.84

Abbreviations: BC, Breast Cancer; CI, Confidence Interval; PRS, Polygenic Risk Score; SD, Standard Deviation; SE, Standard Error.

[a] Women who developed another type of tumour before inclusion in the Rotterdam Study were excluded (N=114)

[b] Development of a second primary breast tumour at least one year after the first primary breast tumour.

Supplementary references

1. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *European journal of epidemiology*. May 4 2020;doi:10.1007/s10654-020-00640-5
2. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
3. ICD-10 Version:2016. 2016. Available from: https://icd.who.int/browse10/2016/en. Accessed March, 2019;

CHAPTER 5

5

The predictive ability of the 313-variant-based polygenic risk score for contralateral breast cancer risk prediction in women of European ancestry with a heterozygote *BRCA1* or *BRCA2* pathogenic variant

Inge M.M. Lakeman[*], Alexandra J. van den Broek[*], Juliën A.M. Vos, Daniel R. Barnes, Julian Adlard, Irene L. Andrulis, Adalgeir Arason, Norbert Arnold, Banu K. Arun, Judith Balmaña, Daniel Barrowdale, Javier Benitez, Ake Borg, Trinidad Caldés, Maria A. Caligo, Wendy K. Chung, Kathleen B.M. Claes, GEMO Study Collaborators, EMBRACE Collaborators, J. Margriet Collée, Fergus J. Couch, Mary B. Daly, Joe Dennis, Mallika Dhawan, Susan M. Domchek, Ros Eeles, Christoph Engel, D. Gareth Evans, Lidia Feliubadaló, Lenka Foretova, Eitan Friedman, Debra Frost, Patricia A. Ganz, Judy Garber, Simon A. Gayther, Anne-Marie Gerdes, Andrew K. Godwin, David E. Goldgar, Eric Hahnen, Christopher R. Hake, Ute Hamann, Frans B.L. Hogervorst, Maartje J. Hooning, John L. Hopper, Peter J. Hulick, Evgeny N. Imyanitov, OCGN Investigators, HEBON Investigators, GEMO Investigators, Claudine Isaacs, Louise Izatt, Anna Jakubowska, Paul A. James, Ramunas Janavicius, Uffe Birk Jensen, Yue Jiao, Esther M. John, Vijai Joseph, Beth Y. Karlan, Carolien M. Kets, Irene Konstantopoulou, Ava Kwong, Clémentine Legrand, Goska Leslie, Fabienne Lesueur, Jennifer T. Loud, Jan Lubiński, Siranoush Manoukian, Lesley McGuffog, Austin Miller, Denise Molina Gomes, Marco Montagna, Emmanuelle Mouret-Fourme, Katherine L. Nathanson, Susan L. Neuhausen, Heli Nevanlinna, Joanne Ngeow Yuen Yie, Edith Olah, Olufunmilayo I. Olopade, Sue K. Park, Michael T. Parsons, Paolo Peterlongo, Marion Piedmonte, Paolo Radice, Johanna Rantala, Gad Rennert, Harvey A. Risch, Rita K. Schmutzler, Priyanka Sharma, Jacques Simard, Christian F. Singer, Zsofia Stadler, Dominique Stoppa-Lyonnet, Christian Sutter, Yen Yen Tan, Manuel R. Teixeira, Soo Hwang Teo, Alex Teulé, Mads Thomassen, Darcy L. Thull, Marc Tischkowitz, Amanda E. Toland, Nadine Tung, Elizabeth J. van Rensburg, Ana Vega, Barbara Wappenschmidt, Peter Devilee, Christi J. van Asperen, Jonine L. Bernstein, Kenneth Offit, Douglas F. Easton, Matti A. Rookus, Georgia Chenevix-Trench, Antonis C. Antoniou[#], Mark Robson[#], and Marjanka K. Schmidt[#]

[*] Contributed equally to this manuscript
[#] Shared last authors

Lakeman IMM, et al. Genetics in Medicine 2021;23:1726-1737

Abstract

Purpose: To evaluate the association between a previously published 313-variant-based breast cancer (BC) polygenic risk score (PRS_{313}) and contralateral breast cancer (CBC) risk, in *BRCA1* and *BRCA2* pathogenic variant heterozygotes.

Methods: We included women of European ancestry with a prevalent first primary invasive BC (*BRCA1*=6,591 with 1,402 prevalent CBC cases; *BRCA2*=4,208 with 647 prevalent CBC cases) from CIMBA, a large international retrospective series. Cox regression analysis was performed to assess the association between overall and ER-specific PRS_{313} and CBC risk.

Results: For *BRCA1* heterozygotes the estrogen receptor (ER)-negative PRS_{313} showed the largest association with CBC risk, HR per SD=1.12, 95%CI [1.06-1.18], C-index=0.53; for *BRCA2* heterozygotes, this was the ER-positive PRS_{313}, HR=1.15, 95%CI [1.07-1.25], C-index=0.57. Adjusting for family history, age at diagnosis, treatment or pathological characteristics for the first BC did not change association effect sizes. For women developing first BC <age 40 years, the cumulative PRS_{313} 5th and 95th percentile 10-year CBC risks were 22% and 32% for *BRCA1* and 13% and 23% for *BRCA2* heterozygotes, respectively.

Conclusion: The PRS_{313} can be used to refine individual CBC risks for *BRCA1/2* heterozygotes of European ancestry, however the PRS_{313} needs to be considered in the context of a multifactorial risk model to evaluate whether it might influence clinical-decision-making.

Introduction

Heterozygotes of germline pathogenic variants in *BRCA1* or *BRCA2* (henceforth: *BRCA1/2* heterozygotes) have a higher risk of developing contralateral breast cancer than non-heterozygotes[1]. The estimated cumulative 10-year contralateral breast cancer risk varies across studies between 18.5%-34.2% for *BRCA1* heterozygotes and between 10.8%-29.2% for *BRCA2* heterozygotes[1-6], compared to 4-6% in the population[7, 8]. Whether or not to undergo a risk-reducing contralateral mastectomy, which is an invasive intervention and associated with side effects such as postoperative surgical complications, inability to breast feed in the future and psychosocial burden[9], is an important and difficult decision for *BRCA1/2* heterozygotes who have been just confronted with their first breast cancer diagnosis. Precise individualized risk estimates could facilitate decision making for these women.

Two important factors influencing contralateral breast cancer risk in *BRCA1/2* heterozygotes are the age at diagnosis of the first breast tumor and a family history of breast cancer[2, 4, 5, 10]. The effect of family history on contralateral breast cancer risk suggests a role for other genetic factors. In the last decade, more than 180 common low risk variants have been associated with breast cancer risk in Genome Wide Association Studies[11-13]. Individually, these variants are associated with small increases in risk, but when combined as polygenic risk scores (PRS) they may improve disease-related risk stratification for women of European and Asian ancestry in the population[14-16]. A limited number of studies have shown that variants associated with the risk of a first primary breast cancer are also associated with the risk of contralateral breast cancer[17-19]. Furthermore, the PRS derived from the general population has also been shown to be associated with breast cancer risk in *BRCA1/2* heterozygotes[20-24].

The most predictive, well validated PRS, for breast cancer in the general population is based on 313 breast cancer-associated variants (PRS_{313}); it showed an association with breast cancer in ten prospective studies with an odds ratio (OR) per standard deviation (SD) of 1.61 and an area under the receiver-operator characteristic curve of 0.630[14]. Among *BRCA2* heterozygotes, this same PRS_{313} was also associated with breast cancer risk, hazard ratio (HR) per SD=1.31, 95%CI [1.27-1.36][24]. Among *BRCA1* heterozygotes, the largest association with breast cancer risk was found using the estrogen receptor (ER)-negative PRS_{313} (which uses the same variants but with weights adapted to provide better prediction for ER-negative disease), HR=1.29, 95%CI [1.25-1.33][24]. Although these effect sizes were smaller than those for the general population, the 313-variant-based PRS could have a substantial impact on the high absolute risks[24], associated with *BRCA1/2* pathogenic variants[25]. Whether variants associated with breast cancer are associated with contralateral breast cancer risk for *BRCA1/2* heterozygotes as well, individually or combined in a PRS,

has not been investigated previously. If so, the PRS may be useful to guide choices for risk management, especially regarding invasive risk-reducing contralateral mastectomy. In this study, we investigated whether the 313-variant-based PRS for breast cancer are associated with contralateral breast cancer risk among women of European ancestry with pathogenic variants in *BRCA1/2* and explored the implications for contralateral breast cancer risk prediction for these women.

Materials and Methods

Study participants

We used retrospective cohort data from heterozygotes participating in the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA)[26]. Briefly, CIMBA participants are heterozygotes of pathogenic variants in *BRCA1* or *BRCA2* who are 18 years or older at the time of inclusion and have phenotypic data available[26]. CIMBA includes eighty-one individual studies of which the majority of the participants were ascertained through cancer genetics clinics[26]. Although studies in CIMBA include individuals of non-European ancestry, our analyses were, due to power considerations (small numbers available for analyses and expected lower estimates for the PRS_{313} in Asian ancestry based on results of women in the general breast cancer population[19]), restricted to women of European ancestry with available array genotyping data (31,195 women of 67 studies).

Women were eligible for this retrospective analysis if they developed an invasive primary breast tumor without metastatic disease at least 1 year before the baseline age. Women without information about metastatic disease were assumed to have no metastatic disease (n=9,242 of whom 2,140 had a known negative lymph node status). Baseline age was defined as the age at local ascertainment (97%), or when this was not known, age at genetic testing (2%) or age at last follow-up (1%). Women were excluded if no information was available about the age at baseline or if they had developed synchronous contralateral breast cancer. Synchronous contralateral breast cancer was defined as contralateral breast cancer within one year after the first primary breast cancer, which was based on the exact date of cancer diagnosis or, if this was not available, on the age at diagnosis. A schematic overview of the selection is shown in Figure S1. In total, 6,591 women with *BRCA1* and 4,208 women with *BRCA2* pathogenic variants were included in this study, among whom 1,402 *BRCA1* heterozygotes and 647 *BRCA2* heterozygotes have had contralateral breast cancer. The diagnosis of primary and contralateral breast cancer was confirmed by pathology records, tumor registry data or medical records by the individual studies. Available phenotypic information for all participants is shown in Table 1, including the number of participants for whom the information was not available for

each of the variables. Information about the ER-status of the first primary breast cancer compared to the contralateral breast cancer is shown in Table S1.

Genotyping and Polygenic Risk Score calculation

For most of the participants, genotyping was performed with the Illumina OncoArray[27]. The remaining participants were genotyped with the Illumina iCOGS array[11]. Details about the quality control procedures and correlation between the arrays have been described previously[19, 24, 28-31]. European ancestry was determined using genetic data and multidimensional scaling. More detailed information about the genotyping and PRS calculation is provided in the supplementary methods.

We used the 313-variant-based PRS for breast cancer developed in an independent study using data from the general population as described previously[14]; correlation between PRS based on the two genotyping arrays was high[19]. The PRS for overall breast cancer (PRS_{313}) and two ER-specific PRS, the ER-positive PRS_{313} and ER-negative PRS_{313} were calculated. The variants and their corresponding weights used in the PRS as published previously[14], and the imputation quality are listed in Table S2. The three PRS were standardized to the mean from all CIMBA participants, including both unaffected and affected women, and to the SD in BCAC population controls which were included in the validation dataset[14]. Using these SDs, the HR estimates for the associations of the standardized PRS_{313} in our study are directly comparable with the OR estimates reported in the BCAC population-based study[14] and the HR estimates reported for primary breast cancer in *BRCA1* and *BRCA2* heterozygotes[24].

Statistical analysis

To assess the associations between the three PRS and contralateral breast cancer risk in *BRCA1/2* heterozygotes, Cox-regression analyses were performed. The time at risk was started one year after the first breast cancer diagnosis based on the exact date or if not available, on the age of developing the first breast tumor. Time at risk of participants was censored at age at baseline, i.e., end of follow-up in these analyses, prophylactic contralateral mastectomy, or death, whichever was earlier (Figure S2). Incidence of a metachronous contralateral breast cancer, invasive or *in situ*, before baseline was considered as an event in the main analyses. The proportional hazard assumption was evaluated by using Schoenfeld residuals against the transformed time. A sensitivity analysis was performed considering invasive contralateral breast cancer only as an event. Women who developed an *in situ* contralateral breast cancer were censored at the age at diagnosis of the *in situ* contralateral breast cancer. Furthermore, a sensitivity analysis was performed including information about distant relapse, which was available for 1,725 *BRCA1* and 1,450 *BRCA2* heterozygotes. In total 55 *BRCA1* heterozygotes and 101 *BRCA2* heterozygotes were censored at the age of distant relapse of which 13 and 11 women

were excluded from the analyses, respectively, because they developed distant relapse in the year before the baseline age.

Analyses were stratified by country (Table S3), adjusted for birth cohort (quartiles of the observed distribution), and clustered on family membership using a unique family-identifier to account for the inclusion of related individuals. For *BRCA1* and *BRCA2* respectively, there were 5923 and 3752 clusters of which 554 and 362 clusters had more than one participant. The main analyses assessed the association with the PRS as a continuous covariate. We evaluated the linearity of the association using restricted cubic splines with three knots, which showed no evidence for violation of the linearity assumption. The discriminatory ability of the best performing PRS was evaluated by Harrell's C-index[32]. C-indexes were calculated stratified by country and clustered on family membership.

The influence of possible confounding variables on the observed associations was assessed using the PRS exhibiting the largest associations. Possible confounding variables included breast cancer family history, age at diagnosis of the first breast cancer, pathological characteristics and treatment of the first breast cancer. Each variable was added to the model one by one and in addition, a full model that included all possible confounders together was fitted. If the addition of a variable resulted in a change of more than 10% in the log HR, the variable was retained as a covariate in the final Cox-regression model. To avoid excluding many participants with missing data for one of these included variables (Table 1), missing data were imputed using Multiple Imputation by Chained Equations (MICE)[33]. Imputation was started with the least missing variable and progressed in order of increased amount of missing data. Using this method, 10 complete data sets for analyses were created and mean parameter estimates were derived.

Secondary analyses were performed for ER-positive and ER-negative cases only, based on the ER-status of the contralateral breast cancer, after imputation as described above. The average number of ER-positive and ER-negative cases in the 10 imputed data sets is shown in Table S4. In these analyses the event of interest was either ER-positive or ER-negative contralateral breast cancer. Contralateral breast cancer cases with the alternative ER-status were censored at the age of contralateral breast cancer.

The interaction between the PRS with the age at first breast cancer diagnosis was tested in the final model, treating the PRS as a continuous variable. Furthermore, the effect size of the PRS was evaluated for groups based on the age at first primary breast cancer diagnosis (<40 years; 40 to 50 years; ≥50 years)[1, 20]. The association of the PRS and contralateral breast cancer risk was tested separately for heterozygotes of pathogenic variants that lead to unstable or no protein (class I) and heterozygotes of pathogenic variants that lead to mutant stable protein (class II). Finally, analyses were performed to test the association

between a categorized PRS and contralateral breast cancer risk to establish whether the results were consistent with those under a continuous PRS model. The categories were defined on the basis of the distribution of the PRS in unilateral breast cancer cases, using PRS percentiles (0-5th, 5th-10th, 10th-20th, 20th-40th, 40th-60th (reference), 60th-80th, 80th-90th, 90th-95th, 95th-100th).

Cumulative risks

Absolute contralateral breast cancer risks were calculated at percentiles of the best-performing continuous PRS for both *BRCA1* and *BRCA2* heterozygotes, using the log HR per SD and including an interaction term with the continuous age at first breast cancer diagnosis (at age 35; 45 and 55 for the corresponding age groups as described below). For this purpose, we constrained the incidence of contralateral breast cancer, by age at first breast cancer and in years after the first breast cancer, and averaged over all PRS categories to agree with external contralateral breast cancer incidence estimates, as described previously[23]. These external incidence estimates were based on prospective cohort data from three consortia on heterozygotes of pathogenic *BRCA1* and *BRCA2* variants[1], the International *BRCA1/2* Carrier Cohort Study (IBCCS), the Breast Cancer Family Registry (BCFR), and the Kathleen Cuningham Foundation Consortium for Research Into Familial Breast Cancer (kConFab). Because the contralateral breast cancer incidences vary with the age of first breast cancer diagnosis, incidences were calculated for three different groups based on the age of the first breast cancer diagnosis (<40 years, 40 to 50 years, ≥50 years)[1].

All statistical tests were performed with R version 3.5.0[34]. Statistical significance was defined as a two-sided p-value <0.05.

Results

In the analyses, 6,591 *BRCA1* and 4,208 *BRCA2* heterozygotes of European ancestry who had developed an invasive first primary breast cancer before entry in CIMBA were identified. The median follow-up time was 6.0 and 5.4 years for *BRCA1* and *BRCA2* heterozygotes, respectively. In total, 1,402 *BRCA1* and 647 *BRCA2* heterozygotes were diagnosed with a metachronous contralateral breast cancer before enrollment in CIMBA. The cumulative 10-year risk of developing contralateral breast cancer in this cohort was 25%, 95%CI [23.5%-26.4%] and 18.8%, 95%CI [17.1%-20.5%] for *BRCA1* and *BRCA2* heterozygotes, respectively (Figure S3). Patient and tumor characteristics as well as the PRS distributions are shown in Table 1 and Figure S4.

Table 1. Characteristics of the participants

		BRCA1 heterozygotes		*BRCA2* heterozygotes	
		UBC, n (%)	CBC, n (%)	UBC, n (%)	CBC, n (%)
N		5,189	1,402	3,561	647
Genotyping Array	**iCOGS**	895 (17)	200 (14)	383 (11)	80 (12)
	OncoArray	4,294 (83)	1,202 (86)	3,178 (89)	567 (88)
Birth cohort	**<1920**	25 (0.5)	8 (0.6)	23 (0.6)	9 (1)
	1920-1929	143 (3)	46 (3)	121 (3)	30 (5)
	1930-1939	392 (8)	130 (9)	341 (10)	99 (15)
	1940-1949	1,060 (20)	386 (28)	793 (22)	172 (27)
	1950-1959	1,540 (30)	452 (32)	1,104 (31)	202 (31)
	1960-1969	1,354 (26)	298 (21)	822 (23)	115 (18)
	≥1970	675 (13)	82 (6)	357 (10)	20 (3)
Variant class[a]	**I**	3,354 (65)	904 (64)	3,207 (90)	570 (88)
	II	1,345 (26)	374 (27)	125 (4)	25 (4)
	III	490 (9)	124 (9)	229 (6)	52 (8)
BRRM		160 (3)	0	101 (3)	0
Deceased	**N**	44 (0.8)	12 (0.9)	19 (0.5)	2 (0.3)
Family history[b]	**No BC**	583 (11)	175 (12)	289 (8)	78 (12)
	1 BC	906 (17)	270 (19)	760 (21)	127 (20)
	≥ 2 BC	1,250 (24)	363 (26)	1,120 (31)	210 (32)
	Unknown	2,450 (47)	594 (42)	1,392 (39)	232 (36)
Characteristics of first BC					
Age at diagnosis	**Mean**	41.8	38.5	44.5	41.8
	Range	19-82	19-68	18-85	21-75
ER status	**Positive**	570 (11)	92 (7)	1,302 (37)	182 (28)
	Negative	1,738 (33)	402 (29)	424 (12)	61 (9)
	Unknown	2,881 (56)	908 (65)	1,835 (52)	404 (62)
Node status	**Positive**	797 (15)	182 (13)	781 (22)	119 (18)
	Negative	1,544 (30)	441 (31)	877 (25)	151 (23)
	Unknown	2,848 (55)	779 56)	1,903 (53)	377 (58)
Tumor size[c]	**T1**	1,261 (24)	314 (22)	842 (24)	136 (21)
	T2	771 (15)	211 (15)	553 (16)	87 (13)
	T3	67 (13)	12 (0.9)	78 (2)	8 (1)
	T4	16 (0.5)	2 (0.1)	22 (0.6)	2 (0.3)
	Unknown	3,074 (59)	863 (62)	2,066 (58)	414 (64)
Chemotherapy[d]	**Yes**	1,099 (21)	236 (17)	821 (23)	123 (19)
	No	576 (11)	212 (15)	503 (14)	129 (20)
	Unknown	3,514 (68)	954 (68)	2,237 (63)	395 (61)
Adjuvant hormone therapy	**Yes**	493 (10)	125 (9)	795 (22)	111 (17)
	No	1,103 (21)	288 (21)	474 (13)	135 (21)
	Unknown	3,593 (69)	989 (71)	2,292 (64)	401 (62)
Adjuvant trastuzumab therapy	**Yes**	11 (0.2)	1 (0.1)	20 (0.6)	0 (0)
	No	1,161 (22)	351 (25)	983 (28)	218 (34)
	Unknown	4,017 (77)	1,050 (75)	2,558 (72)	429 (66)
Radiotherapy	**Yes**	1,090 (21)	277 (20)	797 (22)	158 (24)
	No	535 (10)	141 (10)	420 (12)	84 (13)
	Unknown	3,564 (69)	984 (70)	2,344 (66)	405 (63)

Characteristics of CBC					
Age at diagnosis	**Mean**	-	47.3	-	51.24
	Range	-	26-80.5	-	23.8-86
Invasiveness	**Invasive**	-	1,267 (90)	-	545 (84)
	Non-invasive	-	135 (10)	-	102 (16)
ER-status	**Positive**	-	101 (7)	-	197 (30)
	Negative	-	446 (32)	-	50 (8)
	Unknown	-	855 (61)	-	400 (62)
PRS_{313}					
Standardized PRS_{313} mean (SD)	**Overall BC**	0.08 (1.01)	0.13 (1.01)	0.09 (1.02)	0.27 (1.04)
	ER-positive BC	0.07 (1.01)	0.09 (1.01)	0.08 (1.01)	0.27 (1.03)
	ER-negative BC	0.09 (1.00)	0.23 (0.99)	0.07 (1.02)	0.23 (1.07)

[a] Variant class: I=unstable or no protein, II= stable mutant protein, III= consequence unknown.
[b] Family history was defined as the number of first- or second- degree relatives affected with BC, ranging from 0 to ≥2.
[c] Tumor size: T1=≤2cm (≤0.79in), T2=>2cm-5cm (>0.79-1.97in), T3=>5cm (>1.97in), T4=any size, with direct extension to the chest wall or skin.
[d] Including neoadjuvant and adjuvant chemotherapy
Abbreviations: BC, Breast Cancer; BRRM, Bilateral Risk Reducing Mastectomy; CBC, Contralateral Breast Cancer; ER-status, Estrogen Receptor status of the tumor; N, Number; PRS, Polygenic Risk Score; SD, Standard Deviation; UBC, Unilateral Breast Cancer

PRS and contralateral breast cancer risk

Results of the association analyses between the PRS and contralateral breast cancer risk are shown in Table 2, Table S4 and Figure 1.

BRCA1 heterozygotes

For *BRCA1* heterozygotes the ER-negative PRS_{313} showed the largest association with all contralateral breast cancer, HR per SD=1.12, 95%CI [1.06-1.18], p-value=6.0×10^{-5}, C-index 0.53, 95%CI [0.51-0.55]. There was no evidence of violation of the proportional hazard assumption, p-value=0.840.

Neither sequential inclusion of possible confounders, nor including all these confounders in one model, changed the log HR estimate for the ER-negative PRS_{313} association more than 10% when compared with the model with no confounders (Table S5).

Considering only invasive contralateral breast cancer as the event of interest resulted in a similar association with the ER-negative PRS_{313}, HR per SD=1.13, 95%CI [1.07-1.20], p-value=3.2×10^{-5}.

Censoring at distant metastasis relapse, if applicable, did not change the effect size of the ER-negative PRS_{313}, HR per SD=1.12, 95%CI [1.06-1.18], p-value=$4.9x10^{-5}$.

The HR-estimates for association with contralateral breast cancer for different quantiles of the ER-negative PRS_{313}, were consistent with the predicted HRs from the model using the continuous ER-negative PRS_{313} (Table 2 and Figure 2).

For ER-positive contralateral breast cancer as event, the PRS_{313} showed the largest association, HR per SD=1.32, 95%CI [1.12-1.56], p-value=0.002. For ER-negative contralateral breast cancer as event, only the ER-negative PRS_{313} showed a significant association, HR per SD=1.07, 95%CI [1.01-1.15], p-value=0.036 (Table S4).

BRCA2 heterozygotes

For *BRCA2* heterozygotes the largest association was seen with the ER-positive PRS_{313}, HR per SD=1.15, 95%CI [1.07-1.25], p-value=$1.9x10^{-4}$, C-index 0.57, 95%CI [0.54-0.59]. There was no evidence of violation of the proportional hazard assumption, p-value=0.300.

Neither sequential inclusion of possible confounders, nor including all these confounders in one model, changed the log HR estimate for the ER-positive PRS_{313} association more than 10% when compared with the model with no confounders (Table S5).

Considering only invasive contralateral breast cancer as the event of interest resulted in a similar association, HR per SD for the ER-positive PRS_{313}=1.15, 95%CI [1.06-1.25], p-value=$6.0x10^{-4}$.

Censoring at distant metastasis relapse, if applicable, did not change the effect size of the ER-positive PRS_{313}, HR per SD=1.15, 95%CI [1.07-1.24], p-value=$2.1x10^{-4}$.

The HR estimates for association with contralateral breast cancer for different quantiles of the ER-positive PRS_{313}, were consistent with the predicted estimates using the continuous PRS_{313} (Table 2 and Figure 2).

The ER-positive PRS_{313} showed the largest association with ER-positive contralateral breast cancer for *BRCA2* heterozygotes, HR per SD=1.22, 95%CI [1.11-1.33], p-value=$2.2x10^{-5}$ (Table S4). None of the PRS showed significant associations with ER-negative contralateral breast cancer for *BRCA2* heterozygotes, but the ER-negative PRS_{313} exhibited the largest HR estimate, HR per SD=1.10, 95%CI [0.91-1.32], p-value=0.346.

Table 2: Results of association analyses between the PRS_{313} and contralateral breast cancer risk

		BRCA1 heterozygotes [ER-negative PRS_{313}]					*BRCA2* heterozygotes [ER-positive PRS_{313}]				
		UBC cases, n	CBC cases, n	HR[a]	95% CI	P	UBC cases, n	CBC cases, n	HR[a]	95% CI	P
PRS continuous	**All CBC**	5,189	1,402	1.12	1.06-1.18	$5.98x10^{-5}$	3,561	647	1.15	1.07-1.25	$1.94x10^{-4}$
	Invasive CBC	5,324	1,267	1.13	1.07-1.20	$3.15x10^{-5}$	3,663	545	1.15	1.06-1.25	$6.02x10^{-4}$
Categorical PRS percentiles	**0-5**	260	48	0.81	0.59-1.11	0.188	166	28	1.06	0.71-1.58	0.782
	5-10	259	54	0.77	0.57-1.03	0.082	198	26	0.68	0.44-1.04	0.074
	10-20	519	131	0.94	0.76-1.15	0.544	355	51	0.91	0.66-1.25	0.554
	20-40	1,038	230	0.83	0.70-0.98	0.031	697	108	0.87	0.68-1.13	0.295
	40-60 [reference]	1,037	282	1.00			695	123	1.00		
	60-80	1,038	313	1.04	0.88-1.22	0.664	734	128	0.96	0.75-1.23	0.748
	80-90	519	170	1.11	0.92-1.34	0.255	358	90	1.35	1.03-1.77	0.030
	90-95	259	82	1.18	0.92-1.51	0.185	178	46	1.35	0.96-1.90	0.082
	95-100	260	92	1.24	0.98-1.56	0.074	180	47	1.31	0.94-1.82	0.116
PRS*age BC1 continuous	**Main effect**	5,189	1,402	1.48	1.15-1.89	$2.03x10^{-3}$	3,561	647	1.53	1.11-2.12	0.010
	Interaction effect			0.99	0.99-1.00	0.025			0.99	0.99-1.00	0.089
PRS effect per age group	**<40**	2,339	815	1.22	1.14-1.31	$4.79x10^{-8}$	1,238	268	1.23	1.09-1.38	$5.78x10^{-4}$
	40-50	1,821	456	0.99	0.90-1.09	0.785	1,306	261	1.19	1.05-1.34	$6.91x10^{-3}$
	≥50	1,029	131	1.03	0.86-1.24	0.715	1,017	118	0.97	0.81-1.15	0.698
Variant class[b]	**Class I**	3,354	904	1.11	1.03-1.18	$4.32x10^{-3}$	3,207	570	1.16	1.07-1.26	$1.99x10^{-4}$
	Class II	1,345	374	1.15	1.04-1.28	$4.75x10^{-3}$	125	25	0.91	0.65-1.28	0.594

[a] HRs for association with breast cancer and the continuous PRS_{313} are reported per standard deviation of the PRS in population-based controls.
[b] Class I pathogenic variants result in an unstable or no protein. Class II pathogenic variants yield stable mutant proteins.
Abbreviations: BC1, First primary Breast Cancer; CBC, Contralateral Breast Cancer; CI, Confidence Interval; HR, Hazard Ratio; PRS, Polygenic Risk Score; UBC, Unilateral Breast Cancer.

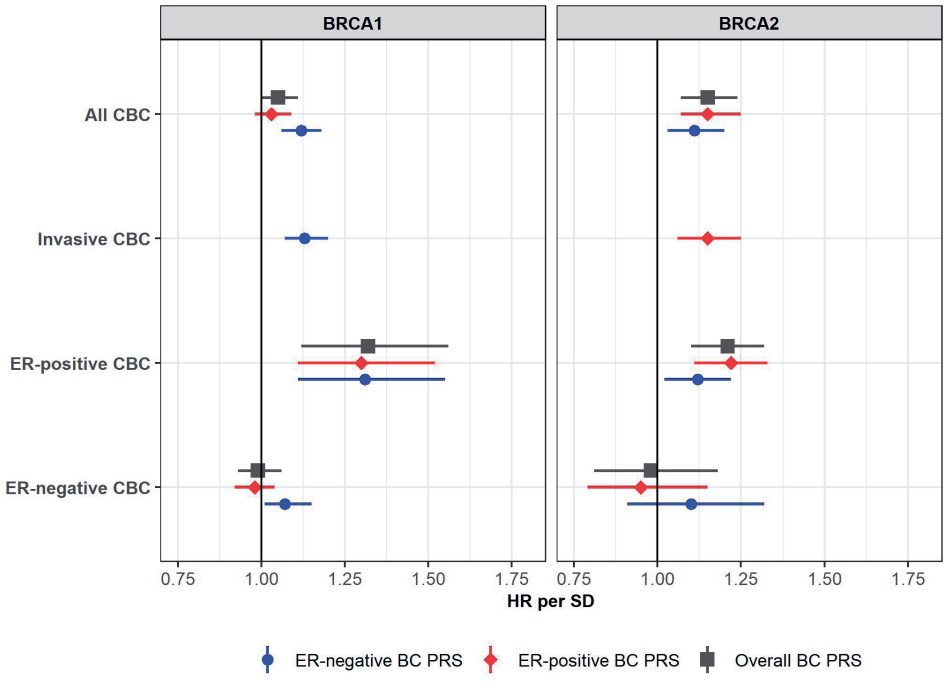


Figure 1: Association between the PRS and contralateral breast cancer risk for BRCA1 and BRCA2 heterozygotes

The figure includes the effect size of the association between contralateral breast cancer and the three different PRS313 after testing for covariates for the following selections: all contralateral breast cancer, invasive contralateral breast cancer only, ER-negative contralateral breast cancer, and ER-positive contralateral breast cancer. The numbers of unilateral and contralateral breast cancer cases and effect sizes are shown in Table 2 and Table S4.

Abbreviations: CBC, Contralateral Breast Cancer; ER, Estrogen Receptor; HR, Hazard Ratio; PRS, Polygenic Risk Score; SD, Standard Deviation.

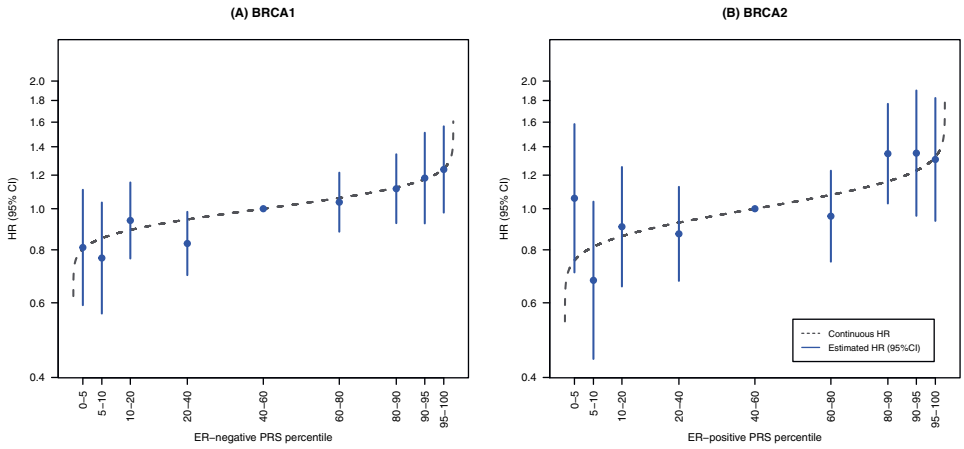


Figure 2: Association between categories of the PRS and contralateral breast cancer risk for BRCA1 and BRCA2 heterozygotes

HRs and 95%CI for percentiles of the ER-negative PRS313 for BRCA1 heterozygotes and the ER-positive PRS313 for BRCA2 heterozygotes, relative to the middle quintile. The PRS percentile groups were 0-5%, 5-10%, 10-20%, 20-40%, 40-60% [reference], 60-80%, 80-90%, 90-95%, and 95-100% based on the distribution in unilateral breast cancer cases. The numbers and corresponding effect sizes are shown in Table 2. The grey line represents the distribution based on the HR of the continuous ER-negative PRS313 and ER-positive PRS313 and the distribution in unilateral breast cancer cases of BRCA1 and BRCA2 heterozygotes respectively.

Abbreviations: CI, Confidence Interval; ER, Estrogen Receptor; HR, Hazard Ratio; PRS, Polygenic Risk Score.

Interaction with age at first breast cancer diagnosis

A significant interaction between the age at first breast cancer diagnosis and the ER-negative PRS_{313} was found for *BRCA1* heterozygotes: HR per year=0.99, 95%CI [0.99-1.00], p-value=0.025. For *BRCA2* heterozygotes a similar magnitude of interaction was observed with the ER-positive PRS_{313}, although the interaction was not significant, HR per year=0.99, 95%CI [0.99-1.00], p-value=0.09.

Categorizing age at first breast cancer diagnosis for *BRCA1* heterozygotes resulted in HRs per SD of the ER-negative PRS_{313} of 1.22, 95%CI [1.14-1.31], 0.99, 95%CI [0.90-1.09] and 1.03, 95%CI [0.86-1.24] for ages <40 years, 40-50 years and ≥50 year respectively. For *BRCA2* heterozygotes the corresponding estimates for ER-positive PRS_{313} were 1.23, 95%CI [1.09-1.38], 1.19, 95%CI [1.05-1.34] and 0.97, 95%CI [0.81-1.15] respectively (Table 2).

Analyses by predicted variant effect on protein expression

For *BRCA1* heterozygotes, the HRs for association between the ER-negative PRS_{313} and contralateral breast cancer risk were similar for heterozygotes of pathogenic variants, which lead to a stable mutant protein (class II) compared with those leading to no protein

or an unstable protein (class I). For *BRCA2* heterozygotes, the ER-positive PRS_{313} effect size for the association with contralateral breast cancer risk was non-significantly smaller among heterozygotes of a pathogenic variant that lead to a stable mutant protein, although statistical power to detect these associations was low and the confidence intervals overlap with the overall estimate (Table 2).

Cumulative risks

Estimate cumulative contralateral breast cancer risks, by categories of age at diagnosis of the first breast cancer are shown in Figure 3. The largest risk difference was seen for women with a first breast cancer diagnosis before the age of 40, with *BRCA1* heterozygotes at the 5th percentile of the ER-negative PRS_{313} having a 10- and 20-year risk of 22% and 35% compared with 32% and 49% at the 95th percentile, respectively. For *BRCA2* heterozygotes, the 10- and 20-year risks in this category were 13% and 25% at the 5th percentile of the ER-positive PRS_{313} compared with 23% and 42% for women at the 95th percentile.

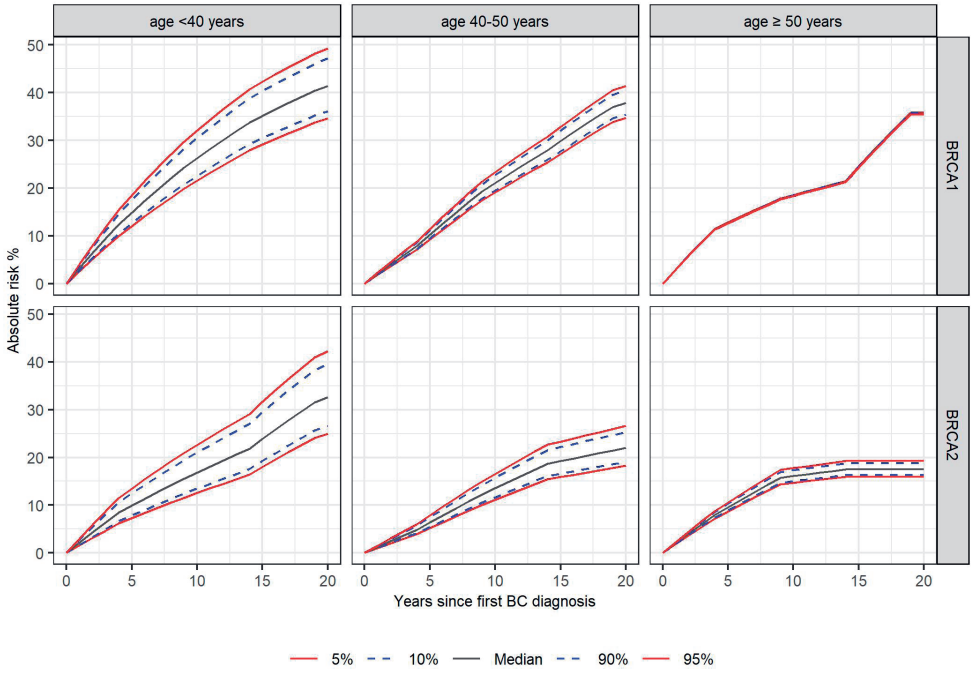


Figure 3: Absolute contralateral breast cancer risk by PRS percentiles per age category of the first breast cancer diagnosis for BRCA1 and BRCA2 heterozygotes

Predicted absolute contralateral breast cancer risks by percentile of the continuous ER-negative PRS313 for BRCA1 heterozygotes and ER-positive PRS313 for BRCA2 heterozygotes. The assumed contralateral breast cancer incidences were from a study that estimated breast cancer incidence in a large prospective cohort of BRCA1 and BRCA2 heterozygotes20. The age categories were based on

the age at diagnosis of the first primary breast tumor. Risks were calculated including the interaction between the PRS and the continuous age of first breast cancer diagnosis. The lines for different percentiles of the PRS are overlapping for the age category ≥50 year for BRCA1 heterozygotes. Abbreviations: BC, Breast Cancer; CBC, Contralateral Breast Cancer; PRS, Polygenic Risk Score.

Discussion

In this study we investigated the associations between an established PRS based on 313 variants for primary first breast cancer and contralateral breast cancer risks among *BRCA1* and *BRCA2* heterozygotes of European ancestry enrolled in the large international retrospective CIMBA cohort. We showed significant albeit modest associations among both *BRCA1* and *BRCA2* heterozygotes between the PRS and contralateral breast cancer risk. For *BRCA1* heterozygotes, the largest association was seen with the ER-negative PRS_{313}, while for *BRCA2* heterozygotes, both the PRS_{313} and ER-positive PRS_{313} showed similar associations with contralateral breast cancer risk that were somewhat larger than the ER-negative PRS_{313} association. These findings are consistent with previous studies on the effects of disease-specific PRS on the first breast cancers in *BRCA1 and BRCA2* heterozygotes[20, 24] and with the higher relative prevalence of ER-negative and ER-positive contralateral breast cancers respectively, in this cohort.

For both *BRCA1* and *BRCA2* heterozygotes, the strength of the association was greater for ER-positive contralateral breast cancers compared with ER-negative contralateral breast cancers (in the case of *BRCA1*, even if the ER-negative PRS was used), although most of the confidence intervals overlapped. The effect sizes for the PRS are also larger for ER-positive disease in the general population, perhaps because ER-positive disease is commoner and the power to identify genetic variants has been greater for ER-positive disease. With larger data sets, it should be possible to develop better subtype specific PRS for contralateral breast cancer.

Although we found clear associations between the PRS and contralateral breast cancer risk, the magnitude of these associations (expressed in terms of HRs) were smaller than previously reported for the first breast cancers. For *BRCA1* heterozygotes, the HR per SD for the association between the ER-negative PRS_{313} and breast cancer was 1.29, 95%CI [1.25-1.33][24], compared with 1.12, 95%CI [1.06-1.18] for contralateral breast cancer in this study. For *BRCA2* heterozygotes, the HR per SD for the association between the ER-positive PRS_{313} and breast cancer was 1.31, 95%CI [1.26-1.36][24], compared with 1.15, 95%CI [1.07-1.24] for contralateral breast cancer in this study. This lower relative risk is consistent with a general pattern of a lower relative risk in a higher risk population, as seen in, the lower relative risk for contralateral breast cancer than first breast cancer in the general population[19], and the lower relative risk for the first cancer in *BRCA1/2* heterozygotes than in the general

population[24]. The attenuated estimate might be explained by several factors, some of which are speculative. *BRCA1/2* pathogenic variant heterozygotes in this study were selected based on having a first breast cancer; these women will have on average a higher PRS, but also higher frequencies of other genetic and non-genetic risk factors than women who do not develop breast cancer at all. This can lead to a weaker association with the PRS as women with the largest PRS may have lower risks due to other factors, a phenomenon related to index event bias[35]. There could also be negative interactions between the PRS effect and other risk factors (for example, treatment factors). However, in this study, we have shown that adjustment for the known contralateral breast cancer risk factors did not change the effect size of the PRS, which was also shown in population-based studies[17, 19]. Finally, although we tried to exclude potential early metastases misdiagnosed as second primaries by excluding women who developed a contralateral breast cancer the first year after the primary diagnosis, it is possible that a small percentage of contralateral breast cancers were metastases[36].

A limitation of this study is that participants were recruited through clinical genetic centers, resulting in ascertainment bias, as individuals are more likely to have a strong family of breast cancer and/or be affected at a young age in order to be referred for testing. This was a historical cohort in which follow-up was prior to entry into CIMBA, so that all cases are prevalent. Therefore, the breast cancer patients included in the analyses are likely to be at higher contralateral breast cancer risk when compared with the general *BRCA1/2* heterozygote breast cancer population. Indeed, the estimated 20-year risks of developing contralateral breast cancer in this study were higher compared to a previously published study with a prospective design[1]: 47% versus 40% for *BRCA1* heterozygotes and 40% versus 26% for *BRCA2* heterozygotes, respectively. While this is unlikely to introduce a significant bias in the relative risk estimates, a prospective cohort would clearly be preferably, although this will take several years to achieve. Finally, the PRS was developed using data sets of women of European ancestry, since our dataset included insufficient samples of women of other ancestries, and our results were exclusively based on women of European ancestry. Therefore, caution is required when applying this to non-European ancestry populations. However, a population study found clear associations between the PRS, based on the same 313 variants or a subset of these variants, and (contralateral) breast cancer also in women of Asian ancestry. The effect size of these associations were slightly weaker, possibly reflecting the fact that this PRS was developed in a cohort of women of European ancestry[16, 19]. These results suggest that there might be an association with the PRS as well in *BRCA1/2* heterozygotes of Asian ancestry. Future studies including a sufficient number of individuals of Asian ancestry are needed to confirm this statement.

Although the relative risks of the PRS for contralateral breast cancer were modest, differences in the PRS may still have an important effect on the absolute risk, which is

high. *BRCA1* and *BRCA2* heterozygotes under age 40 at first breast cancer, at the 5th and 95th percentile of the PRS differed by 10% in 10-year contralateral breast cancer risk. These absolute risk differences are modest, but might be of relevance for the choices regarding preventive surgery if incorporated into a multifactorial model that includes other predictive factors, such as family history and adjuvant systemic treatment of the first breast cancer[37, 38]. In the context of such a comprehensive model, further research is needed to investigate whether the PRS would contribute to the choices that women make for follow-up or preventive surgery.

To summarize, we have investigated the associations between PRS based on 313 variants with contralateral breast cancer risk in a large international series of *BRCA1/2* heterozygotes. We found that the PRS is associated with contralateral breast cancer risk in both *BRCA1* and *BRCA2* heterozygotes of European ancestry and that PRS can be used to refine estimates of contralateral breast cancer risks in these women. However, for women with a first breast cancer after the age of 50, PRS may be of less value in the prediction of the contralateral breast cancer risk. Incorporating risk factors other than PRS and including ER-specific estimates may further improve contralateral breast cancer risk prediction. Before implementation in a diagnostic setting, our results should be validated in a prospective cohort of *BRCA1* and *BRCA2* heterozygotes.

Acknowledgements and funding

This work was supported by the Alpe d'HuZes/Dutch Cancer Society (KWF Kankerbestrijding) project 6253 and Dutch Cancer Society (KWF Kankerbestrijding) project UL2014-7473. See online version for further details.

We acknowledge all the families, clinicians, family doctors, researchers, research nurses, research assistants, and technicians who contribute to the individual studies of which we used the data for this research and manuscript. See online version for further details.

Ethics Statement

All participants were recruited by the host institutions under protocols approved by local ethics review boards and provided written informed consent[24].

Disclosure of potential conflicts of interest

Claudine Isaacs is consultant to Astra Zeneca, Novartis, Pfizer, Genentech, PUMA, Seattle Genetics and received research support from Tesaro.

Data availability statement

The CIMBA data is available on request. To receive access to the data, a concept form must be submitted, which will then be reviewed by the CIMBA Data Access Coordination Committee (DACC). Please contact Lesley McGuffog (e-mail: ljm26@medschl.cam.ac.uk), to get access to these concept forms (http://cimba.ccge.medschl.cam.ac.uk/contact/).

References

1. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *Jama*. Jun 20 2017;317(23):2402-2416. doi:10.1001/jama.2017.7112
2. Graeser MK, Engel C, Rhiem K, et al. Contralateral breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Dec 10 2009;27(35):5887-92. doi:10.1200/jco.2008.19.9430
3. Mavaddat N, Peock S, Frost D, et al. Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. *Journal of the National Cancer Institute*. Jun 5 2013;105(11):812-22. doi:10.1093/jnci/djt095
4. Metcalfe K, Gershman S, Lynch HT, et al. Predictors of contralateral breast cancer in BRCA1 and BRCA2 mutation carriers. *British journal of cancer*. Apr 26 2011;104(9):1384-92. doi:10.1038/bjc.2011.120
5. Rhiem K, Engel C, Graeser M, et al. The risk of contralateral breast cancer in patients from BRCA1/2 negative high risk families as compared to patients from BRCA1 or BRCA2 positive families: a retrospective cohort study. *Breast cancer research : BCR*. Dec 7 2012;14(6):R156. doi:10.1186/bcr3369
6. van der Kolk DM, de Bock GH, Leegte BK, et al. Penetrance of breast cancer, ovarian cancer and contralateral breast cancer in BRCA1 and BRCA2 families: high cancer incidence at older age. *Breast cancer research and treatment*. Dec 2010;124(3):643-51. doi:10.1007/s10549-010-0805-3
7. Lizarraga IM, Sugg SL, Weigel RJ, Scott-Conner CE. Review of risk factors for the development of contralateral breast cancer. *American journal of surgery*. Nov 2013;206(5):704-8. doi:10.1016/j.amjsurg.2013.08.002
8. Kramer I, Schaapveld M, Oldenburg HSA, et al. The Influence of Adjuvant Systemic Regimens on Contralateral Breast Cancer Risk and Receptor Subtype. *Journal of the National Cancer Institute*. Jul 1 2019;111(7):709-718. doi:10.1093/jnci/djz010
9. Carbine NE, Lostumbo L, Wallace J, Ko H. Risk-reducing mastectomy for the prevention of primary breast cancer. *The Cochrane database of systematic reviews*. Apr 5 2018;4:Cd002748. doi:10.1002/14651858.CD002748.pub4
10. van den Broek AJ, van 't Veer LJ, Hooning MJ, et al. Impact of Age at Primary Breast Cancer on Contralateral Breast Cancer Risk in BRCA1/2 Mutation Carriers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Feb 10 2016;34(5):409-18. doi:10.1200/jco.2015.62.3942
11. 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]
12. 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

13. 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
14. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
15. 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]
16. Ho WK, Tan MM, Mavaddat N, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun*. Jul 31 2020;11(1):3833. doi:10.1038/s41467-020-17680-w
17. 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
18. 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]
19. Kramer I, Hooning MJ, Mavaddat N, et al. Breast Cancer Polygenic Risk Score and Contralateral Breast Cancer Risk. *American journal of human genetics*. Nov 5 2020;107(5):837-848. doi:10.1016/j.ajhg.2020.09.001
20. 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
21. Antoniou AC, Sinilnikova OM, McGuffog L, et al. Common variants in LSP1, 2q35 and 8q24 and breast cancer risk for BRCA1 and BRCA2 mutation carriers. *Human molecular genetics*. Nov 15 2009;18(22):4442-56. doi:10.1093/hmg/ddp372
22. Antoniou AC, Spurdle AB, Sinilnikova OM, et al. Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *American journal of human genetics*. Apr 2008;82(4):937-48. doi:10.1016/j.ajhg.2008.02.008
23. Antoniou AC, Beesley J, McGuffog L, et al. Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction. *Cancer research*. Dec 1 2010;70(23):9742-54. doi:10.1158/0008-5472.Can-10-1907
24. Barnes DR, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in Medicine*. 2020/10/01 2020;22(10):1653-1666. doi:10.1038/s41436-020-0862-x
25. Gail MH, Pfeiffer RM. Breast Cancer Risk Model Requirements for Counseling, Prevention, and Screening. *Journal of the National Cancer Institute*. Sep 1 2018;110(9):994-1002. doi:10.1093/jnci/djy013

26. Chenevix-Trench G, Milne RL, Antoniou AC, Couch FJ, Easton DF, Goldgar DE. An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). *Breast cancer research : BCR*. 2007;9(2):104. doi:10.1186/bcr1670
27. Amos CI, Dennis J, Wang Z, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. Jan 2017;26(1):126-135. doi:10.1158/1055-9965.epi-16-0106
28. Gaudet MM, Kuchenbaecker KB, Vijai J, et al. Identification of a BRCA2-specific modifier locus at 6p24 related to breast cancer risk. *PLoS genetics*. 2013;9(3):e1003173. doi:10.1371/journal.pgen.1003173
29. Couch FJ, Wang X, McGuffog L, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS genetics*. 2013;9(3):e1003212. doi:10.1371/journal.pgen.1003212
30. Kuchenbaecker KB, Neuhausen SL, Robson M, et al. Associations of common breast cancer susceptibility alleles with risk of breast cancer subtypes in BRCA1 and BRCA2 mutation carriers. *Breast cancer research : BCR*. Dec 31 2014;16(6):3416. doi:10.1186/s13058-014-0492-9
31. 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
32. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*. Feb 28 1996;15(4):361-87. doi:10.1002/(sici)1097-0258(19960229)15:4<361::Aid-sim168>3.0.Co;2-4
33. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*. Mar 2011;20(1):40-9. doi:10.1002/mpr.329
34. R_Core_Team_(2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
35. Dahabreh IJ, Kent DM. Index Event Bias as an Explanation for the Paradoxes of Recurrence Risk Research. *Jama*. 2011;305(8):822-823. doi:10.1001/jama.2011.163 %J JAMA
36. Begg CB, Ostrovnaya I, Geyer FC, et al. Contralateral breast cancers: Independent cancers or metastases? *International journal of cancer*. Jan 15 2018;142(2):347-356. doi:10.1002/ijc.31051
37. Akdeniz D, Schmidt MK, Seynaeve CM, et al. Risk factors for metachronous contralateral breast cancer: A systematic review and meta-analysis. *Breast (Edinburgh, Scotland)*. Apr 2019;44:1-14. doi:10.1016/j.breast.2018.11.005
38. Giardiello D, Steyerberg EW, Hauptmann M, et al. Prediction and clinical utility of a contralateral breast cancer risk model. *Breast cancer research : BCR*. Dec 17 2019;21(1):144. doi:10.1186/s13058-019-1221-1

Supplementary figures and tables

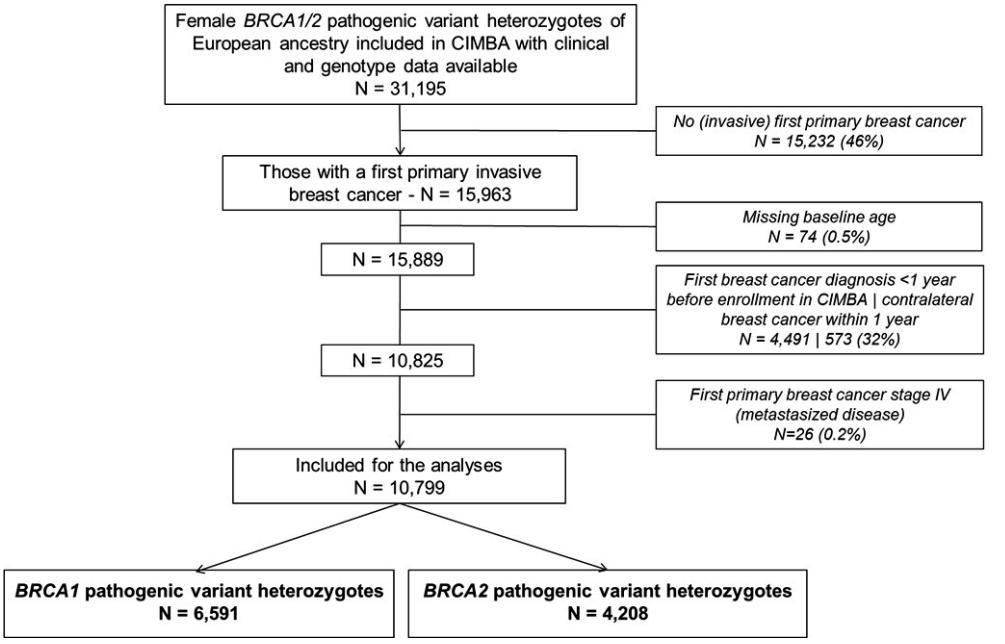


Figure S1: Flow chart of the inclusion of CIMBA participants
Flow chart of the inclusion and exclusion of CIMBA participants for this study.
Abbreviation: N, Number

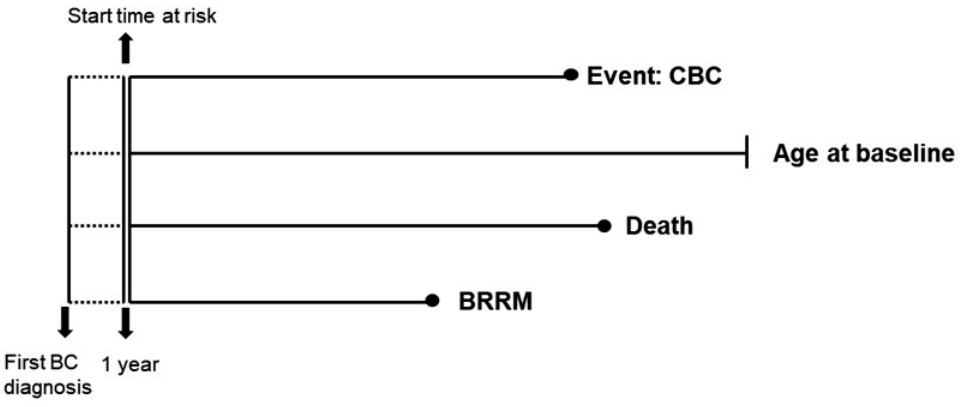


Figure S2: Time at risk in the association analyses
The time at risk was assumed to start one year after the first breast cancer. Participants were censored at (i) age at baseline, (ii) bilateral risk reducing mastectomy or (iii) death, whichever was earlier. Baseline age was defined as the age at local ascertainment (97%), or when this was not known, age at genetic testing (2%) or age at last follow-up (1%). Incidence of a metachronous contralateral breast cancer, invasive or *in situ*, before baseline was considered as an event in the main analyses.
Abbreviations: BC, Breast Cancer; BRRM, Bilateral Risk Reducing Mastectomy; CBC, Contralateral Breast Cancer.

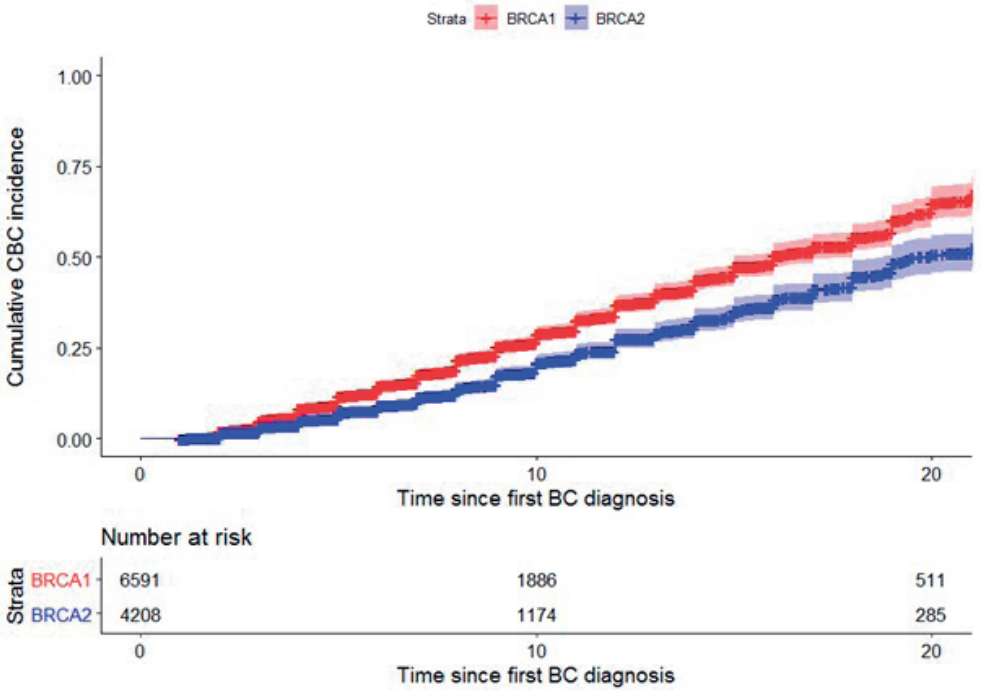


Figure S3: Cumulative contralateral breast cancer incidence for *BRCA1* and *BRCA2* heterozygotes since the first breast cancer diagnosis

Plot of the cumulative contralateral breast cancer incidence for *BRCA1* (red) and *BRCA2* (blue) pathogenic variant heterozygotes. Confidence intervals are shown with the transparent red and blue color. The time of follow-up started at the age of first primary invasive breast cancer diagnosis. Abbreviations: BC, Breast Cancer; CBC, Contralateral Breast Cancer.

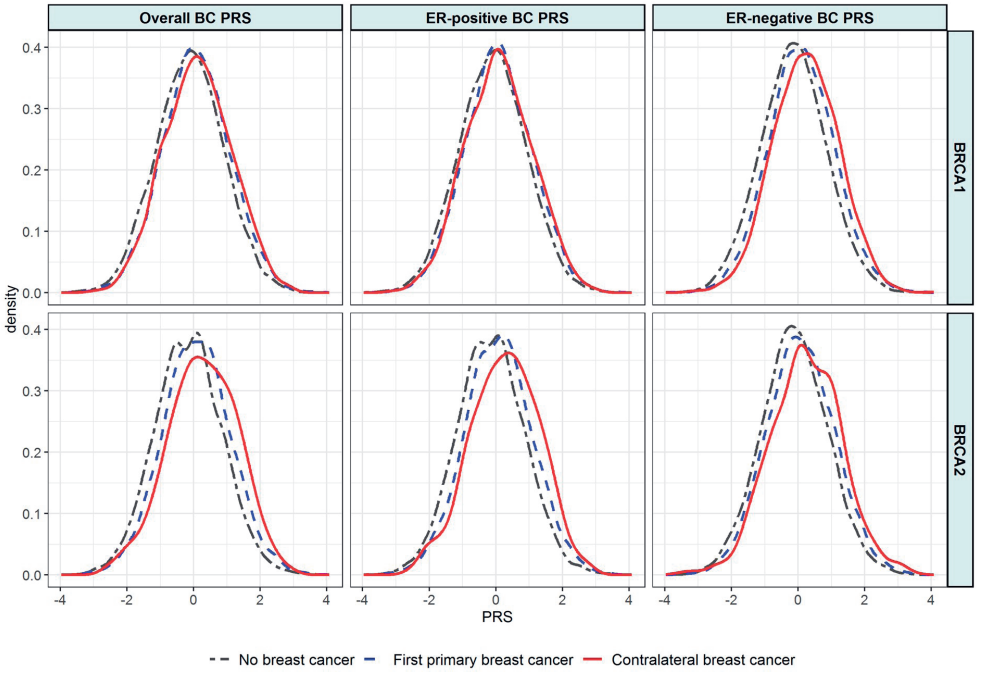


Figure S4: Distribution of the overall breast cancer, ER-positive and ER-negative PRS_{313} for *BRCA1* and *BRCA2* heterozygotes without breast cancer, with a first primary breast cancer and with contralateral breast cancer

Density plots of the standardized PRS distributions for *BRCA1* and *BRCA2* heterozygotes. The distributions are shown for CIMBA participants who did not develop breast cancer (grey two-dashed line), who developed an invasive first primary breast cancer only (blue dashed line, selection shown in Figure S1) and who developed a metachronous contralateral breast cancer (red solid line). The number of included women for these groups were 8,837, 5,189, and 1,402 for *BRCA1* heterozygotes and 5,665, 3,561, and 647 for *BRCA2* heterozygotes.

Abbreviations: BC, Breast Cancer; ER, Estrogen Receptor; PRS, Polygenic Risk Score.

Table S1: Estrogen receptor status of the first primary breast tumor and the contralateral breast tumor

	ER-status BC1	ER-status CBC		
		ER-positive	ER-negative	Unknown
***BRCA1* heterozygotes**	**ER-positive**	25	42	25
	ER-negative	29	256	117
	Unknown	47	148	713
***BRCA2* heterozygotes**	**ER-positive**	100	19	63
	ER-negative	16	18	27
	Unknown	81	13	310

Abbreviations: BC1, first primary Breast Cancer; CBC, Contralateral Breast Cancer; ER, Estrogen Receptor.

Table S2: 313 variants included in the polygenic risk score
See online material. First nine columns of the table were published by Mavaddat et al.[1]

Table S3: Country of origin of included CIMBA participants

Country of origin		*BRCA1* heterozygotes	*BRCA2* heterozygotes
Group[a]	**Country**		
Africa	**South Africa**	29	70
America	**Brazil**	0	1
	Canada	209	103
	United States of America	1266	735
Asia	**Israel**	60	52
	Qatar	0	1
Australia	**Australia**	355	269
Eastern Europe	**Albania**	1	0
	Czech Republic	41	0
	Hungary	120	36
	Latvia	9	0
	Lithuania	62	6
	Poland	217	0
	Russia	12	0
Northwestern Europe	**Austria**	179	77
	Belgium	128	43
	Denmark	224	171
	Ireland	1	1
	Finland	46	44
	France	677	565
	Germany	762	394
	Iceland	0	102
	Netherlands	440	196
	Sweden	177	24
	United Kingdom	702	614
Southern Europe	**Greece**	99	13
	Italy	472	285
	Portugal	23	58
	Spain	280	348

[a] Groups for country used in the cox-regression analyses

Table S4: Results of the association analyses between the PRS and contralateral breast cancer risk

Outcome	PRS_{313}	*BRCA1* heterozygotes					*BRCA2* heterozygotes				
		UBC cases, n	CBC cases, n	HR	95% CI	P	UBC cases, n	CBC cases, n	HR	95% CI	P
All CBC	**Overall BC**	5,189	1,402	1.05	1.00 -1.11	0.059	3,561	647	1.15	1.07-1.24	$2.33x10^{-4}$
	ER-positive			1.03	0.98-1.09	0.208			1.15	1.07-1.25	$1.94x10^{-4}$
	ER-negative			1.12	1.06-1.18	$5.98x10^{-5}$			1.11	1.03-1.20	0.005
ER-positive CBC	**Overall BC**	6,312[a]	279[a]	1.32	1.12-1.56	0.002	3,701[a]	507[a]	1.21	1.10-1.32	$4.19x10^{-5}$
	ER-positive			1.30	1.11-1.52	0.002			1.22	1.11-1.33	$2.15x10^{-5}$
	ER-negative			1.31	1.11-1.55	0.003			1.12	1.02-1.22	0.014
ER-negative CBC	**Overall BC**	5,468[a]	1123[a]	0.99	0.93-1.06	0.859	4,068[a]	140[a]	0.98	0.81-1.18	0.809
	ER-positive			0.98	0.92-1.04	0.491			0.95	0.79-1.15	0.628
	ER-negative			1.07	1.01-1.15	0.036			1.10	0.91-1.32	0.346

[a] Average number over 10 imputed datasets

Abbreviations: BC, Breast Cancer; CBC, Contralateral Breast Cancer; CI, Confidence Interval; ER, Estrogen Receptor; HR, Hazard Ratio; PRS, Polygenic Risk Score; UBC, Unilateral Breast Cancer.

Table S5: Results of the change in effect size of the association between the PRS and contralateral breast cancer risk, using multivariable Cox Regression models

	Added variable	*BRCA1* heterozygotes; ER-negative PRS_{313}					*BRCA2* heterozygotes; ER-positive PRS_{313}				
		β[a]	% change	HR[a]	95% CI	p	β[b]	% change	HR[b]	95% CI	P
Base model[c]		0.111	ref	1.12	1.06-1.18	$5.98x10^{-5}$	0.143	ref	1.15	1.07-1.25	$1.94x10^{-4}$
	Family history	0.112	1.10	1.12	1.06-1.18	$4.43x10^{-5}$	0.143	0.26	1.15	1.07-1.25	$2.53x10^{-4}$
	Age of BC1	0.112	1.03	1.12	1.06-1.18	$4.32x10^{-5}$	0.151	5.01	1.16	1.08-1.26	$1.29x10^{-4}$
Tumor characteristics BC1	**ER-status**	0.111	0.04	1.12	1.06-1.18	$4.28x10^{-5}$	0.141	1.68	1.15	1.07-1.24	$3.73x10^{-4}$
	Node status	0.112	0.69	1.12	1.06-1.18	$4.65x10^{-5}$	0.145	1.27	1.16	1.07-1.25	$2.21x10^{-4}$
	Tumor size	0.111	0.01	1.12	1.06-1.18	$5.36x10^{-5}$	0.147	2.24	1.16	1.07-1.25	$1.95x10^{-4}$
Therapy BC1	**Chemotherapy**	0.110	0.70	1.12	1.06-1.18	$5.97x10^{-5}$	0.143	0.04	1.15	1.07-1.25	$2.53x10^{-4}$
	Hormone	0.111	0.10	1.12	1.06-1.18	$5.15x10^{-5}$	0.144	0.14	1.15	1.07-1.25	$2.48x10^{-4}$
	Trastuzumab	0.111	0.02	1.12	1.06-1.18	$5.22x10^{-5}$	0.143	0.23	1.15	1.07-1.25	$2.57x10^{-4}$
	Radiotherapy	0.111	0.09	1.12	1.06-1.18	$5.29x10^{-5}$	0.143	0.18	1.15	1.07-1.25	$2.56x10^{-4}$
Full model	**All above variables combined**	0.114	2.24	1.12	1.07-1.18	$4.50x10^{-5}$	0.150	4.37	1.16	1.07-1.26	$2.06x10^{-4}$

[a] Effect size of the ER-negative PRS_{313}
[b] Effect size of the ER-positive PRS_{313}
[c] Cox regression model for the association between the PRS and contralateral breast cancer, stratified by country, clustered on family membership, and adjusted for birth cohort (quartiles of the observed distribution).
Abbreviations: BC1, first primary Breast Cancer; CI, Confidence Interval; HR, Hazard Ratio; PRS, Polygenic Risk Score

Supplementary references

1. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002

CHAPTER 6

6

Comprehensive breast cancer risk prediction for women from non-*BRCA1/2* breast cancer families – an observational pilot study in one Dutch medical centre

I.M.M. Lakeman[1, 2], D.J. Jenner[3], N. van der Stoep[1], F. Baas[1], E. Hahnen[4], A. Brédart[5, 6], A. Tüchler[4], R. Schmutzler[4], E.M.A. Bleiker[1, 3, 7], M. K. Schmidt[1, 3], P. Devilee[2, 8], C. J. van Asperen[1]

This chapter is a draft of a Dutch pilot study that will be published in a full manuscript together with data from France and Germany.

Abstract

Introduction: Our aim was to determine the clinical and emotional impact of using and communicating Comprehensive Risk Prediction (CRP) compared to standard family history-based risk prediction (FHRP).

Methods: In this observational pilot study, we included 38 unaffected first-degree female relatives of women affected with breast cancer, who underwent breast cancer counselling in 2019/2020 and tested negative for pathogenic variants in *BRCA1/2, PALB2, CHEK2*, and *ATM*. During that consultation, the counselee had received a single risk score for their healthy relatives based on FHRP (clinical advice). Individual FHRP and CRP were (re-) calculated by using the CanRisk web tool. CRP included family history, the PRS_{313}, and lifestyle/hormonal factors. CRP results were communicated to the participants via web consultation on individual basis. To assess the psychosocial impact, participants were asked to fill in questionnaires before and after risk communication.

Results: Based on their individual CRP, ten participants changed to a lower, and eight to a higher risk category compared to FHRP. Notably, two sisters who had been given the same FHRP-based moderate risk category, changed respectively to a higher and lower risk category after CRP, mainly due to the PRS_{313}. Moreover, individual FHRP re-calculated with CanRisk differed from the risk category and corresponding clinical management given during the first genetic consultation of the affected family member for 13 out of 38 participants. Participants were overall positive about receiving their CRP, explanation during the web-consultation and method of communication (online versus hospital visit)

Conclusion: In this pilot-sample, 47% of healthy relatives shifted to another risk category and received a different screening advice based on their CRP as compared to their FHRP. The dissimilarity between the initial clinical advice and CanRisk-based FHRP emphasizes the need for standardised tools and protocols.

Introduction

Women with a first-degree relative affected with breast cancer have a twofold increased risk of developing the disease themselves[1]. Over half of the familial risk of breast cancer has been clarified genetically, with rare pathogenic mutations in moderate- and high-risk genes, such as *BRCA1* and *BRCA2*, accounting for ~25%, and common low risk variants associated with breast cancer for a further ~36%[2, 3]. Summarized in a Polygenic Risk Score (PRS), these common low risk variants are useful to stratify women into different risk categories[3-8]. Breast cancer surveillance for unaffected women from breast cancer families is currently guided by risk assessment based on family history and DNA testing results of five breast cancer genes (i.e. *BRCA1/2, PALB2, CHEK2, ATM*). We have shown previously that addition of the polygenic risk score (PRS) to this routine changed screening recommendations for a substantial proportion of the women according to breast screening guidelines[5, 9]. Because secondary prevention by mammogram to reduce the burden of the disease has several disadvantages as well, including overdiagnosis[10], it would be optimal to target those women most likely to benefit from screening by their individual breast cancer risk.

Individual breast cancer lifetime risks can be calculated by various risk prediction algorithms[11], such as the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA)[12]. BOADICEA calculates cumulative breast cancer risk based on family history, mammographic density, lifestyle/hormonal and genetic risk factors, including the most predictive PRS, based on 313 variants (PRS_{313})[3, 12]. This model has been externally prospectively validated[13-16], is implemented in the user friendly CanRisk online tool[17], and has received CE-marking. Although it seems ready for implementation into breast cancer prevention programs, Comprehensive Risk Prediction (CRP) is currently not used in clinical management.

At this moment, genetic testing is mainly offered to women affected by breast cancer and is mainly restricted to the high penetrant genes *BRCA1, BRCA2, PALB2* and the moderate risk genes *CHEK2* and *ATM*. With the possibility for individualised risk prediction (CRP), a new group of unaffected relatives of breast cancer patients become eligible for counselling. Because we know that counselling can be a cause for a wide range of psychosocial problems[18], we should be cautious with this new form of risk prediction. With the current study, we aim to determine the clinical and emotional impact of CRP by measuring how often these unaffected women shift in risk category compared to standard family history-based risk prediction, as well as the psychosocial impact of CRP by measuring cancer worries of counselees after having been given their individual CRP score.

Methods

This study is known as the IBR-study (Individualised Breast cancer Risk prediction study), a pilot observational cohort study at the Leiden University Medical Center (LUMC) which has been approved by the medical ethical committee (NL68501.058.18). The IBR-study is still ongoing in close collaboration with the BRIDGES (Breast cancer RIsk after Diagnostic GEne Sequencing) study[19]. The aim of the BRIDGES study is to build a knowledge base that will allow identification of women at high-risk of breast cancer, in particular through comprehensive evaluation of DNA variants in known and suspected breast cancer genes[20]

Study cohort

The cohort consist of unaffected female relatives from counselees affected with breast cancer. These women were included in 2019/2020 via the outpatient clinic of the Department of Clinical Genetics at the LUMC in Leiden, the Netherlands. After a counselee with breast cancer had tested negative for (likely) pathogenic variants in the breast cancer genes *BRCA1, BRCA2, PALB2, CHEK2 and ATM*, her unaffected first-degree female relatives aged 35-60 years, were invited to participate in the study via an appendix to the family letter. The family letter is part of the diagnostic routine in counselling, in which the healthy relative receives a family history-based clinical management advice for breast screening (henceforth termed "clinical advice"). Women interested to participate in the study were asked to enrol in the Hereditary Breast and Ovarian cancer study in the Netherlands (HEBON)[21], during which they gave informed consent. The HEBON study (initiated in 1999) is an ongoing nationwide cohort study with members from breast cancer families, which arranges prospective follow up through record-linkage with the nationwide cancer and pathology registries. Informed consent for the IBR study was received from 45 participants. An overview of our study flow scheme is shown in Figure 1.

Comprehensive risk prediction

CRP was calculated with the CanRisk webtool in which BOADICEA is implemented. Participants received a saliva sample package at home to collect DNA. Breast cancer genes were tested by a multigene panel of which 5 genes were analysed (*BRCA1, BRCA2, PALB2, ATM* and *CHEK2).* The 313 common low risk variants[3] were genotyped by a slightly modified panel of 340 variants (27 backup variants). Participants were asked to fill in the HEBON questionnaire, including questions about lifestyle/hormonal factors. Four different calculations were performed in the CanRisk webtool.

- FHRP: Family history-based risk prediction including pedigree based family history, and gene panel results of the index and participant
- Non-Genetic Risk Factors (NGRF): FHRP including hormonal/lifestyle risk factors
- PRS_{313}: FHRP including PRS_{313}

- CRP: Full model, i.e. FHRP, NGRF, and the PRS_{313}.

For all four types of calculations, we have reported the 5-year, 10-year and lifetime risk (between age 20 and age 80) for developing breast cancer. Hormonal/lifestyle risk factors included age at menarche, age at menopause, number of children, age at first life birth if applicable, Body Mass Index, height, oral contraception use, and alcohol use.

Risk communication

A web consult was scheduled with the investigators (IMML) or (CJvA) and the participant to communicate the individual breast cancer lifetime risk (CRP) including 10-year risk and corresponding clinical management advice, in comparison with the previous reported risk and corresponding clinical management advice given in the family letter (clinical advice). When a participant shifted to a higher risk category, clinical management was advised as recommended in that risk category. When a participant shifted to a lower risk category, the clinical management advice did not change relative to the clinical advice received in the family letter.

Psychosocial questionnaires

To assess the psychosocial impact, participants were asked to fill in questionnaires before and after communication of the individual breast cancer risk score. Approximately three months before the web consult (T1) participants were asked to fill in two online questionnaires: the Psychosocial Aspects of Hereditary Cancer questionnaire (PAHC)[22], including the Distress thermometer (DT)[23] and the Cancer Worry Scale (CWS)[24] questionnaire. Two months after the web consult (T2), the participant received again the PAHC including the DT and CWS questionnaire. Six months after the web consult (T3), participants were asked to fill in a questionnaire about the uptake of the clinical management advice and experience with counselling.

Descriptive analyses

Summary statistics are shown for all four types of calculations. For all risk calculations, the corresponding risk category was determined based on the Dutch breast cancer screening guideline (Table 1)[25]. The number of individuals who shift to another risk category based on their CRP as compared to family history-based risk prediction (FHRP) was determined. Furthermore, FHRP calculated by CanRisk was compared with the clinical management advice given during counselling of the index (clinical advice).

The psychosocial impact of comprehensive risk prediction (CRP) for unaffected relatives of affected counselees at T1 and T2 will be analysed in the context of the BRIDGES study and will be published by Bredart et al. (*manuscript in preparation*).

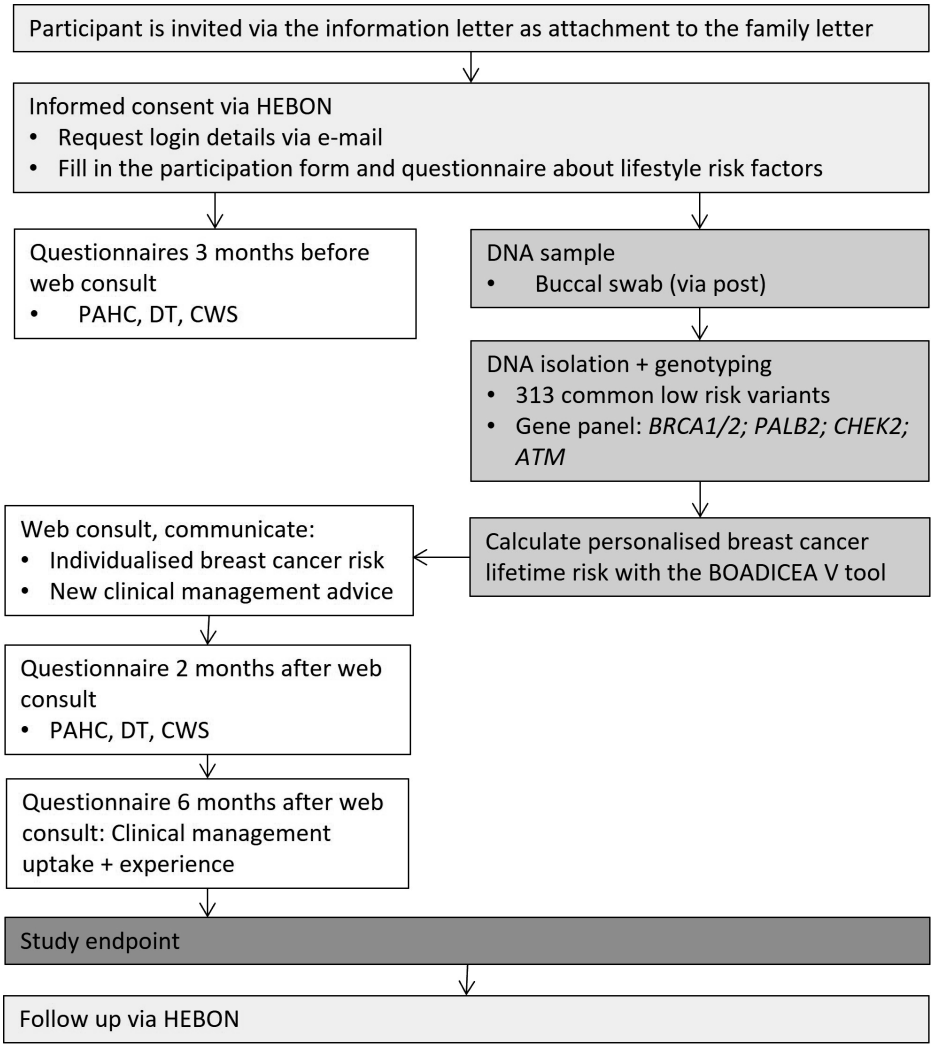


Figure 1. Flow scheme of the IBR study
Abbreviations: CWS, Cancer Worry Scale; DT, Distress Thermometer; PAHC, psychosocial aspects of Hereditary Cancer.

Table 1: Breast cancer screening recommendation in the Netherlands based on lifetime risk of developing breast cancer[25].

	Low (RR<2)	Moderate (RR: 2-3)	High (RR: >3)
Lifetime risk	<20%	20-30%	>30%
Start screening	50yr	40yr	35yr
Mammography	Population screening	<50yr annual >50yr population screening	<60yr annual >60yr population screening
MRI	-	-	-

Results

In total, 45 participants were included in the IBR-study for whom we were able to perform the CRP using the *CanRisk* tool for 38 of these participants (Figure 2). The mean age at inclusion was 45 years with an age range from 35 to 59. All included participants derived from 32 families; 6 families had 2 participants included and the remaining families 1 participant.

The mean difference in lifetime risk of including risk factors, PRS_{313} or both (full model) to FHRP was respectively 2.5%, 4.5% and 5.0% (Table 2). For 18, 24 and 24 participants the risk difference was negative (lower) and for 20, 14, and 14 participants the risk difference was positive (higher). The absolute difference in risk was larger by including their PRS_{313} compared to their risk factors, but the largest when both were included (Figure 3).

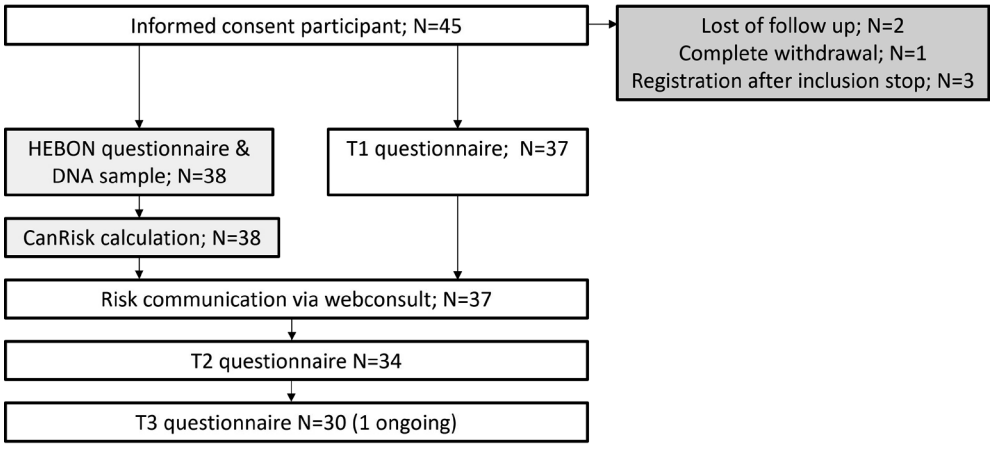


Figure 2. Inclusion of participants in the IBR study
Of the 45 included participants, we were able to calculate breast cancer lifetime risks for 38 of the participants. These risks were only communicated to the participants if they filled in the first psychosocial questionnaires (N=37). For one participant the risk communication was less than two months ago, therefore she has not received the third questionnaire yet.
Abbreviations: N, Number; T, Timepoint.

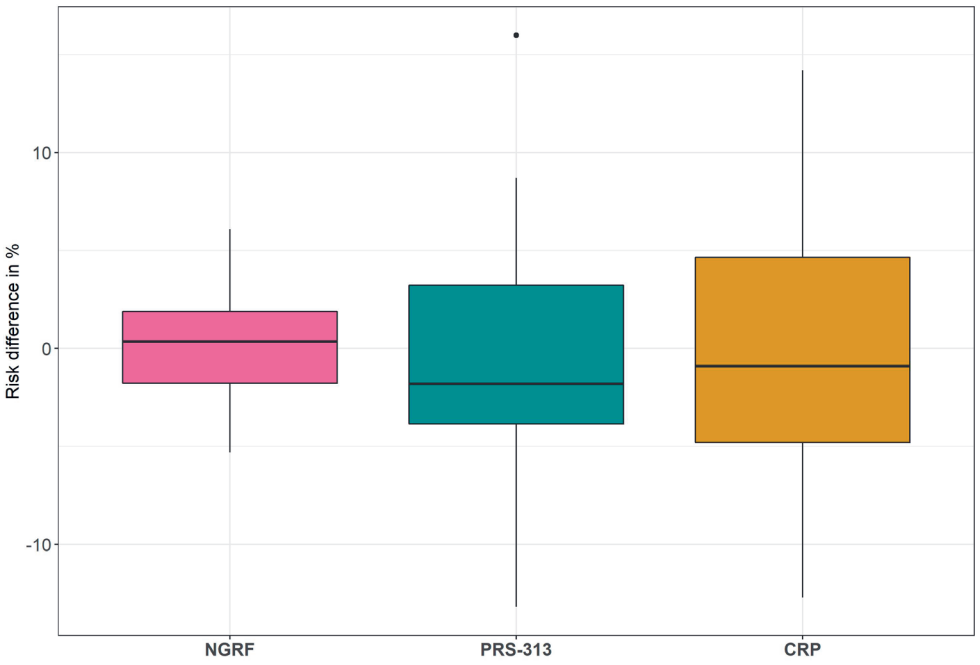


Figure 3: Difference in breast cancer lifetime risk calculated by CanRisk
Boxplot of difference in breast cancer lifetime risk compared to FHRP after a) including hormonal/lifestyle risk factors (purple), b) including the individual PRS_{313} (green), and c) after including both (orange).
Abbreviation: CRP: Comprehensive Risk Prediction (Full model, i.e. FHRP, NGRF, and the PRS_{313}); FHRP, Family History-based Risk Prediction; NGRF, Non-Genetic Risk Factors; PRS, Polygenic Risk Score

Table 2. Difference in lifetime risk compared to FHRP based on 38 participants

	Mean	Lowest	Highest
NGRF	2.5%	0.1%	6.1%
PRS_{313}	4.5%	0.1%	16%
CRP	5.0%	0.1%	14.2%

Abbreviation: CRP, Comprehensive Risk Prediction (Full model, i.e. FHRP, NGRF, and the PRS_{313}); FHRP, Family History-based Risk Prediction; NGRF, Non-Genetic Risk Factors; PRS, Polygenic Risk Score

FHRP calculated by CanRisk versus clinical advice

For 13 out of 38 (34%) of the participants, the risk category based on family history only calculated with *CanRisk* was not consistent with the risk category and corresponding clinical advice given during the genetic consultation of the affected family member. For 12 participants the clinical advice category was higher and for 1 participant it was lower (Table 3).

Table 3. Family -based breast cancer lifetime risk estimated in the clinic versus estimation by CanRisk for 38 healthy women

		FHRP calculated by CanRisk		
		Low	**Moderate**	**High**
Clinical advice	**Low**	7	-	-
	Moderate	6	17	1
	High	-	6	1

Abbreviation: FHRP, Family History based Risk Prediction

CRP versus FHRP calculated by CanRisk

Based on full CRP including both risk factors and PRS_{313}, 10 participants changed to a lower and eight participants to a higher risk category, compared to FHRP calculated by CanRisk (Table 4). Interestingly, two sisters with the same moderate risk category based on their family history, changed respectively to a higher and lower risk category based on their CRP, which was mainly due to their difference in the PRS_{313} (Table 5).

Experiences with CRP

Participants were overall positive about the individual breast cancer risk prediction, explanation during the web-consultation and method of communication (online versus hospital visit) (Figure 4).

Table 4: Family-based risk prediction in CanRisk vs comprehensive risk prediction in CanRisk for 38 healthy women

		CRP calculated by CanRisk		
		Low	**Moderate**	**High**
FHRP calculated by CanRisk	**Low**	10	2	1
	Moderate	9	9	5
	High	-	1	1

Abbreviations: CRP, Comprehensive Risk Prediction (Full model, i.e. FHRP, NGRF, and the PRS_{313}); FHRP, Family History-based Risk Prediction; NGRF, Non-Genetic Risk Factors; PRS, Polygenic Risk Score.

Table 5: Breast cancer lifetime risk scores for 12/38 participants with a family member included

Family	Individual	Relation	Clinical advice	Lifetime risk percentage			
				FHRP	NGRF	PRS_{313}	CRP
1	**1**	2nd degree (aunt/niece)	Moderate	20.8	22.7	20.7	22.6
	2			19.4	18.4	19.9	18.8
2	**1**	1st degree (sisters)	Moderate	20.2	26.0	20.0	25.8
	2			20.2	21.7	26.3	28.2
3	**1**	1st degree (sisters)	Low	16.1	14.3	15.7	13.9
	2			16.2	13.6	13.7	11.4
4	**1**	1st degree (sisters)	High	24.1	22.6	32.8	31.1
	2			24.3	24.9	16.9	17.3
5	**1**	1st degree (sisters)	Moderate	13.7	13.1	9.8	9.3
	2			13.6	10.9	15.9	12.7
6	**1**	1st degree (sisters)	Moderate	24.5	22.8	27.2	25.4
	2			24.7	19.4	22.8	17.7

[a]Small risk differences between sisters are due to birth year difference.
Abbreviations: CRP, Comprehensive Risk Prediction (Full model, i.e. FHRP, NGRF, and the PRS_{313}); FHRP, Family History-based Risk Prediction; NGRF, Non-Genetic Risk Factors; PRS, Polygenic Risk Score

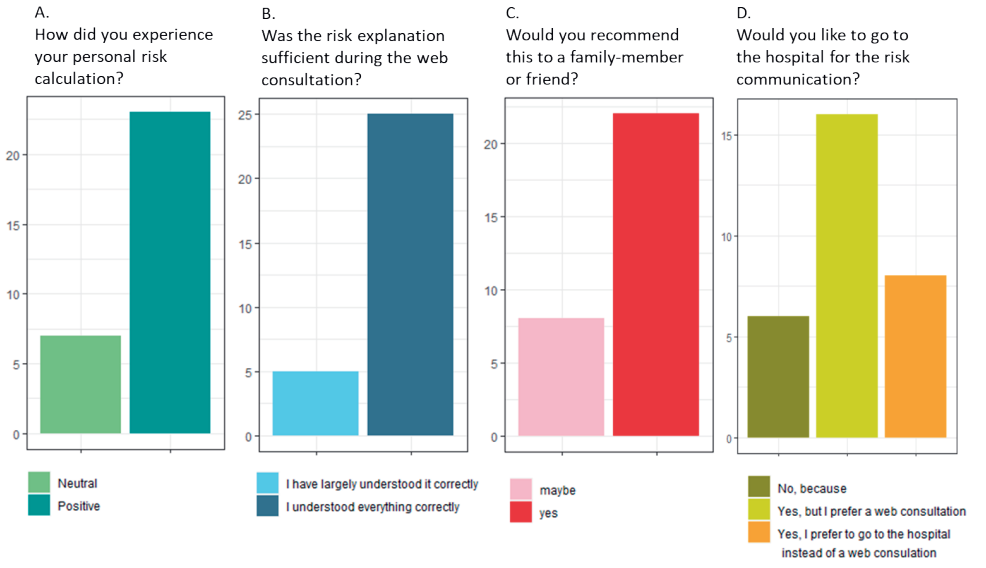


Figure 4: Experience of participants with their individual breast cancer risk prediction and communication via a web consultation

Discussion

This small single-centre pilot study illustrates the potential clinical impact of using CRP in the clinic for healthy relatives of counselees affected with breast cancer. In our study, 18 out of 38 women (47%) shifted to another risk category and received another screening advice based on their CRP calculated by the CanRisk tool as compared to the current standard risk prediction including only family history. Although this small numbers are not statistically significant, this is substantially higher than found in our previous analyses of high-risk research families[5] and clinic-based moderate-risk families[9]. The difference in percentage may be caused by including only unaffected women in comparison to a mixed group of cases and healthy relatives[5] or cases only[9]. However, the number of included participants in our pilot study was too low to draw conclusions from this comparison.

The risk category based on family history only calculated with CanRisk was not always consistent with the clinical advice given during the genetic consultation of the affected family member. Although they are both based on family history only, the risk category was different for 34% of the participants. The main reason for this dissimilarity is probably the lack of uniformity of risk prediction in the clinic. Different risk prediction models[26] (e.g. BOADICEA, Tyrer-Cuzick, Claus) are used in clinical genetic services in the Netherlands for breast cancer risk prediction to guide clinical management for healthy relatives from breast cancer families. Furthermore, clinical management will sometimes be chosen based on clinical view, for example if the predicted lifetime risk is close to a risk category cut of point (i.e. 20% or 30%, Table 1). It would improve consistency if a single risk prediction algorithm is used in the clinic, such as CanRisk.

Psychosocial correlates and details on counselees' and clinical geneticist's perception of CRP from the larger multicenter study of BRIDGES will be presented by Tüchler et al. and Brédart et al. (*manuscripts in preparation)*.

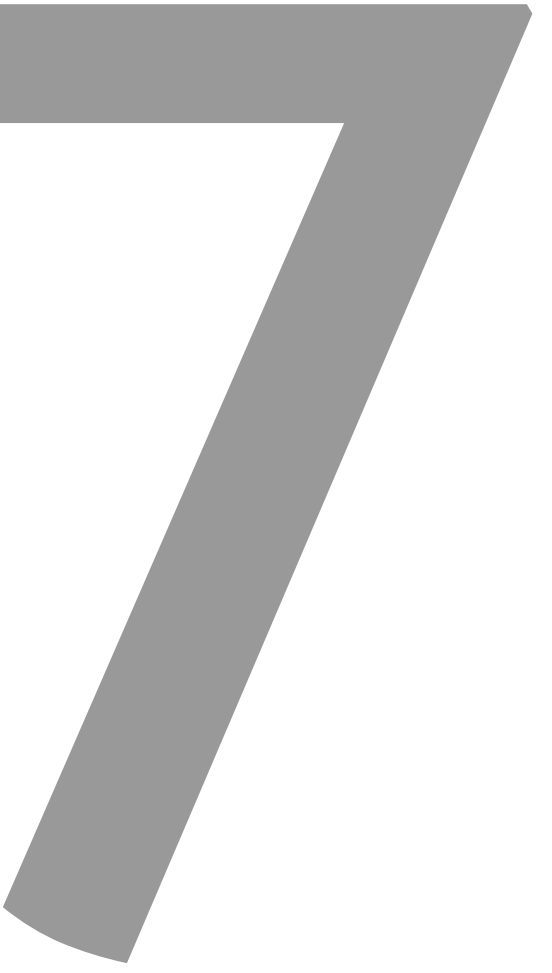
To conclude, we have used CRP in clinical practice on individual level and shown that CRP can shift a substantial proportion of counselees from gene-panel negative breast cancer families to another risk category with consequences for clinical management advice. Furthermore, the dissimilarity between clinical advice based on 'family history only' or based on the Canrisk-calculation emphasizes the need for standardized tools, protocols and training for clinicians.

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. 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. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
4. 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]
5. Lakeman IMM, Hilbers FS, Rodriguez-Girondo M, et al. 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. *Journal of medical genetics*. Sep 2019;56(9):581-589. doi:10.1136/jmedgenet-2019-106072
6. 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]
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. 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
9. Lakeman I, Rodriguez-Girondo M, Lee A, et al. Clinical applicability of the Polygenic Risk Score for breast cancer risk prediction in familial cases. *Submitted to Journal of Medical Genetics*. 2022;
10. Ripping TM, Verbeek AL, Fracheboud J, de Koning HJ, van Ravesteyn NT, Broeders MJ. Overdiagnosis by mammographic screening for breast cancer studied in birth cohorts in The Netherlands. *International journal of cancer*. Aug 15 2015;137(4):921-9. doi:10.1002/ijc.29452
11. 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
12. 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
13. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. *The Lancet Oncology*. Apr 2019;20(4):504-517. doi:10.1016/s1470-2045(18)30902-1

14. Pal Choudhury P, Brook MN, Hurson AN, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. *Breast Cancer Research*. 2021/02/15 2021;23(1):22. doi:10.1186/s13058-021-01399-7
15. Li SX, Milne RL, Nguyen-Dumont T, et al. Prospective Evaluation over 15 Years of Six Breast Cancer Risk Models. *Cancers*. 2021;13(20):5194.
16. Lakeman IMM, Rodríguez-Girondo M, Lee A, et al. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. *Genetics in medicine : official journal of the American College of Medical Genetics*. Nov 2020;22(11):1803-1811. doi:10.1038/s41436-020-0884-4
17. Carver T, Hartley S, Lee A, et al. CanRisk Tool—A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer Epidemiology Biomarkers & Prevention*. 2021;30(3):469-473. doi:10.1158/1055-9965.Epi-20-1319
18. Eijzenga W, Bleiker EM, Hahn DE, Van der Kolk LE, Sidharta GN, Aaronson NK. Prevalence and detection of psychosocial problems in cancer genetic counseling. *Familial cancer*. Dec 2015;14(4):629-36. doi:10.1007/s10689-015-9809-9
19. BRIDGES. Breast Cancer Risk after Diagnostik Gene Sequencing. bridges-research.eu
20. Dorling L, Carvalho S, Allen J, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *The New England journal of medicine*. Jan 20 2021;doi:10.1056/NEJMoa1913948
21. van Asperen CJ, Brohet RM, Meijers-Heijboer EJ, et al. Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. *Journal of medical genetics*. Sep 2005;42(9):711-9. doi:10.1136/jmg.2004.028829
22. Eijzenga W, Bleiker EM, Hahn DE, et al. Psychosocial aspects of hereditary cancer (PAHC) questionnaire: development and testing of a screening questionnaire for use in clinical cancer genetics. *Psycho-oncology*. Aug 2014;23(8):862-9. doi:10.1002/pon.3485
23. Donovan KA, Grassi L, McGinty HL, Jacobsen PB. Validation of the distress thermometer worldwide: state of the science. *Psycho-oncology*. Mar 2014;23(3):241-50. doi:10.1002/pon.3430
24. Custers JA, van den Berg SW, van Laarhoven HW, Bleiker EM, Gielissen MF, Prins JB. The Cancer Worry Scale: detecting fear of recurrence in breast cancer survivors. *Cancer nursing*. Jan-Feb 2014;37(1):E44-50. doi:10.1097/NCC.0b013e3182813a17
25. IKNL. Richtlijn Borstkanker - Screening buiten het bevolkingsonderzoek. Accessed 03-12-2021, https://richtlijnendatabase.nl/richtlijn/borstkanker/screening/screening_buiten_het_bob/screening_buiten_het_bevolkingsonderzoek.html
26. Kim G, Bahl M. Assessing Risk of Breast Cancer: A Review of Risk Prediction Models. *J Breast Imaging*. Mar-Apr 2021;3(2):144-155. doi:10.1093/jbi/wbab001

CHAPTER 7



Discussion and future perspectives

Today, in Dutch clinical genetic services, breast cancer risk prediction is mainly based on family history and carrier status of pathogenic variants in one of the five well known breast cancer genes (i.e. *BRCA1, BRCA2, PALB2, CHEK2*, *ATM*). Family history is an important risk factor for breast cancer. On average, healthy women with at least one first degree relative affected with breast cancer have a relative risk of developing breast cancer of ~2-fold. Last decade, we have gained more knowledge about the aetiology of this familial relative risk, which could improve breast cancer risk prediction in terms of precision and accuracy. Combining all known genetic, familial and lifestyle risk factors will give a more individual based lifetime risk score. In this thesis we have explored the clinical utility of the use of the currently known common low risk variants associated with breast cancer, which explain ~18% of the familial relative risk, for individual breast cancer risk prediction. Especially for families that visit the clinical genetic services in the Netherlands.

7.1 Dutch breast cancer families

In **chapters 2 and 3**, we explored the clinical applicability of the Polygenic Risk Score (PRS) for risk prediction in a cohort of breast cancer families not explained by *BRCA1* or *BRCA2* pathogenic variants, that had visited the clinical genetic services in the Netherlands. It was known that the PRS could improve the discriminative power between breast cancer cases and controls[1-5], but little was known about this discriminative power within families and the additive impact on family-based risk prediction in these families.

In **chapter 2**, high risk breast cancer families were analysed that were selected for genetic research purposes and counselled between 1990 and 2012. An advantage of this cohort was the availability of a DNA sample of both affected and unaffected family members. While most studies use population controls as a reference group[2-4, 6], we used healthy relatives of breast cancer cases as a reference to make it more compatible with clinical practice in clinical genetic services. Only three previous studies have also genotyped breast cancer cases and their unaffected relatives, but with a lower number of variants included in their PRS[7-9]. The PRS in this study was based on 161 breast cancer associated variants which were known at that time[10]. Within our cohort of high-risk families, affected family members had on average a higher PRS compared to their healthy relatives, suggesting already an association between this PRS and breast cancer. Association analyses proved the effect of the PRS, showing a significant association (HR per SD=1.16) within high-risk families between the PRS and breast cancer. As presented in **chapter 3** and described in the literature as well[2, 7], we observed just a very weak positive correlation between the PRS and the family history score, calculated by BOADICEA version 3 using the complete pedigree. This result underscores the additive value of measuring the PRS for every individual in the family, as opposed to using an estimated PRS based on the family history.

With BOADICEA version 3, in which the PRS was not yet implemented[11], lifetime risks (i.e. breast cancer risk between age 20 and age 80) were calculated with and without the PRS in addition to family history-based risk prediction. By adding the PRS, about 20% of both affected and unaffected women were reclassified to another risk category and would have received a different screening advise based on the Dutch breast cancer screening guideline[12].

In **chapter 3** we selected breast cancer cases with a positive family history for breast cancer that visited one of the clinical genetic services in the Netherlands, without a pathogenic variant in *BRCA1* or *BRCA2*. These cases were more representative of the average breast cancer families counselled in the clinic than those analysed in **chapter 2**. The best predictive PRS for breast cancer known at this moment was calculated based on 313 common low risk variants (PRS_{313}). Again, as expected, this PRS was on average higher for breast cancer cases versus population controls. Furthermore, women who developed an *in situ* carcinoma had on average a lower PRS_{313} compared to women who developed an invasive tumour but a higher PRS compared to population controls. Between family members, 13% of the variance in the PRS_{313} could be explained by the PRS_{313} of the proband (case with the youngest diagnosis), hence the proband's PRS was only modestly predictive of that of family members. A significant association was determined in this family-based cohort between breast cancer and the PRS_{313}, OR per SD=1.97, with a stronger effect for invasive compared with *in situ* carcinoma (OR per SD=2.00, and 1.69 respectively). For the majority, gene panel sequencing was performed for at least *CHEK2, ATM* and *PALB2*. In total 1.8% of the controls and 8.4% of the cases carried a truncating pathogenic variant in one of these genes, most frequently in *CHEK2*. Using BOADICEA version 5 where the PRS_{313} is implemented[13], family history-based breast cancer lifetime risk scores were calculated including the pedigree and gene-panel result. In addition to this family history-based score, the individual PRS_{313} was included. For up to 34% of the gene-panel negative cases, screening recommendations could have changed by adding the PRS_{313} to family history-based risk prediction. Addition of the PRS_{313} had a large impact on screening recommendations for *ATM* and *CHEK2* pathogenic variant carriers as well, corresponding to the suggested polygenic effect of moderate risk breast cancer genes. No change was detected for carriers of a *PALB2* pathogenic variant, who all remained in the high-risk category, although variations in risk scores may have impact on choices that women make regarding prophylactic surgery.

These family-based studies are important for implementation of the PRS in the clinic. Using information from breast cancer families which recently visited clinical genetic services, provides a good representation of the group of counselees from families that are seen in the context of clinical genetic services. Furthermore, an advantage of selecting "genetically enriched" cases is that we had a sufficient number of pathogenic variant

carriers in *CHEK2*, and *ATM* in our cohort to show the reclassification (i.e. change to a different screening category) for this group of women as well. However, selecting families with an average higher risk for developing breast cancer, resulting in a higher prevalence of breast cancer in this group compared to the population, causes ascertainment bias so that the effect sizes obtained in these studies cannot be translated directly in the clinic. The higher effect size in our study (**chapter 3,** OR=1.97) compared to population based cohorts of the Breast Cancer Association Consortium (BCAC) (OR=1.61)[14] and in the Dutch population (**chapter 4**, HR=1.56) possibly reflects a higher genetic predisposition in our families. This is also supported by the on average higher PRS for healthy relatives of breast cancer cases compared to population controls and the lower association effect size of the PRS and breast cancer within high-risk families (**chapter 2**, HR=1.16). Although we adjusted for family history, it does probably not suffice to correct for ascertainment bias. This illustrates an important problem in family history-based studies: they lead to overestimation of disease penetrance, which underscores the need for careful separation of family history and the PRS and estimating their effects for the general population. Although we are seeing this selected group of families in the clinic as well, separation of the two risk factors, family history and PRS, will be more specific for an individual. Separation of these risk becomes more important since, compared to 10 years ago, fewer affected families are counselled at this moment.

Both studies showed a quite large percentage of women who changed to another risk category (reclassification) and would have received a different corresponding clinical advice after including the PRS in addition to family history-based risk prediction. These reclassifications were based on breast cancer lifetime risk scores which were mainly calculated for cases (affected counselees), assuming they were 1 year old and unaffected, while in clinical practice the risk scores are only calculated for unaffected family members. Therefore, the reclassification percentage may be different for healthy relatives of affected counselees; the majority of those will have a higher family history-based score, because the affected proband will be included as affected family member. Some studies address this inconsistency by calculating the score for an additional imaginary healthy sister but because the PRS is an individual score, this is not possible in our studies.

In **chapter 6**, we have performed a small pilot study in which we calculated, by using BOADICEA version 5, similar breast cancer lifetime risk scores for 38 unaffected first-degree relatives of women affected with breast cancer, who had already visited the clinical genetic service for breast cancer counselling and tested negative for germline pathogenic variants in one of five breast cancer predisposing genes (*BRCA1/2, PALB2, CHEK2, ATM*). By including family history, non-genetic risk factors, and the PRS, 18 women (47%) changed to a different screening category as compared to the current standard risk prediction including family history alone [Tüchler et al. *manuscript in preparation*]. These results

suggest that if we introduced bias by including cases only, the true reclassification rate for unaffected relatives is probably not lower as the 34% described in **chapter 3**, keeping the conclusion that our results underscore the utility of including the PRS.

Although we found a large percentage of reclassification, the question remains if the direction is correct. Ideally you want to have a prospective cohort of women with data about their screening uptake, breast cancer development and detection of the tumour, i.e. screen detected or interval carcinoma. Unfortunately this information was lacking in both of our cohorts. We had information about the age of diagnosis that would help determine if the cases were retrospectively placed in the right risk category. However, without having information on the detection of the tumour, it is difficult to interpret which category would be the right one. For example, if a woman who was reclassified into the high risk group was diagnosed with breast cancer at age 56, a mammogram biannually via population screening could have been sufficient to detect it, but the recommended annual mammogram for this risk group following the Dutch guideline[12] might have detected the tumour earlier. Based on the knowledge that the BOADICEA model is well calibrated and validated in different prospective studies[15-17] as well as in our study described in **chapter 4** for the Dutch population, we assume that the reclassification leads to the detection of more breast cancers overall and less side-effects of screening such as false positives and overdiagnosis. However, to optimise these benefits of individual risk-based screening, we need to be confident enough to downgrade screening for a part of the women. But even if we can demonstrate cost-efficiency and accept that the reclassification will on average be better for the total group, it remains difficult to translate it to a specific person as seen by a clinician. As clinician you have to decide for that person at that point in time, which method will best manage the real risk for a person and downgrading may be a challenge.

7.2 Breast cancer in the Dutch population

In **chapter 4**, the performance of BOADICEA version 5 and the association with the PRS_{313} was evaluated for the Dutch population. Furthermore, we illustrated the potential impact of the model in detecting breast cancer in a population screening setting in which women would participate based on their individual risk. Comprehensive risk prediction is possible with BOADICEA version 5, which incorporates the PRS_{313} as well as lifestyle, reproductive and hormonal risk factors, but this was not yet validated in the general Dutch population.

We used a large prospective population-based cohort of women aged 45 years or older with extensive follow-up data of up to 25 years. Women who developed breast cancer during follow up had on average a higher PRS_{313} compared to unaffected women. Furthermore, as seen in **chapter 3** as well, women who developed an invasive breast tumour had on

average a higher PRS_{313} compared to women with an *in situ* breast tumour. The PRS_{313} was significantly associated with breast cancer, with a similar effect size (HR=1.56) as in other prospective series of different geographic origin[14], demonstrating its robustness and potential application to the Dutch population. Similar as described in **chapter 3**, the PRS_{313} was associated with *in situ* breast cancer as well, with a non-significant lower effect size than for invasive breast cancer. Moreover, as described previously for a PRS based on 72 variants[18], the PRS_{313} is specifically associated with breast cancer risk and not with a higher risk for the development for a non-breast carcinoma (HR=1.05, non-significant). As determined in previous studies performed by BCAC[4, 14], we found that the effect size of the PRS declined with increasing age. With the BOADICEA model, cumulative 10-year breast cancer risk scores were calculated using four sets of variables (age; age and PRS; age and risk factors; age, PRS, and risk factors). Above inclusion of age, The PRS_{313} improved the discriminatory ability from 0.531 to 0.636. As expected, based on previous research[13, 19], this could only be marginally improved further (to 0.653) by adding lifestyle, reproductive factors, and anthropometric data. Irrespective of the variables included, BOADICEA underestimated the observed risk of 4.4% especially in the highest risk categories. This underestimation was possibly due to the lack of family history data, mammographic density and information about pathogenic variants in *BRCA1/2*. Overall, the PRS_{313} replicates robustly in the Dutch population and the discriminative power of the BOADICEA model seems appropriate for implementation into breast cancer prevention programs. However, for accurate use of the BOADICEA model in the population, information about family history could be important to add.

We illustrated the potential impact of the BOADICEA model in detecting breast cancer in a population screening setting in which women would participate based on their individual risk. In this scenario, the PRS_{313} alone would have detected more cases than the full BOADICEA model (80% versus 62% respectively), but would also have identified a larger screening group (65% versus 45% of all women). Ideally one would want to find the optimal cost-benefit ratio with the highest detection of breast cancer and the lowest false-positive and overdiagnosis rate. An important question in breast cancer risk prediction is how to include and treat *in situ* carcinomas. Although PRS development studies have so far included only invasive breast cancer[4, 14], we showed in **chapter 3** that the PRS_{313} is associated with *in situ* breast cancer as well, consistent with previous research[20] . However, there was a non-significantly lower effect size for *in situ* carcinomas compared to invasive breast cancer. Preferably, comprehensive risk prediction including the PRS_{313} will lead to a higher detection rate of *in situ* carcinomas that are more prone to become invasive and less detection of *in situ* carcinoma that will never become clinically relevant. For this goal, more knowledge is needed about prognostic markers that distinguish between these types. Previous research showed that besides growth pattern, histological grade of ductal carcinoma in situ (DCIS) has been associated with subsequent development of

invasive disease[21-23]. In our study, all women with grade 3 DCIS were in the group eligible for screening based on the PRS and age model and only 50% of the women with a grade 1/2 DCIS. Although the absolute numbers were low, this supports the notion that the PRS predicts DCIS that is more prone to become invasive breast cancer. However, further research needs to be performed to confirm this.

The high prevalence of *in situ* carcinoma nowadays, ~25% of all breast cancers[24], leads to the question, relevant for both family-members and their counsellors, whether women who develop these breast cancers should be considered as "affected" or "unaffected" in family-based risk prediction. For example, BOADICEA is presented as a model that predicts invasive breast cancer considering only invasive breast tumours in the family[13]. Ideally, it would be possible to include DCIS as well in these risk prediction models. However, epidemiological studies determining the risk for developing breast cancer for a healthy relative of someone with DCIS are lacking. Although probably the majority of DCIS will remain indolent[23], it may be possible that DCIS in some individuals within breast cancer families is more prone to become invasive due to genetic predisposition. Therefore, not including DCIS may lead to an underestimation of breast cancer risk in these families. In my opinion, until we are able to distinguish a clinically relevant DCIS from benign DCIS (i.e., overdiagnosis) or until the associated familial relative risk is known and incorporated, we have to include DCIS as invasive breast cancer in risk prediction models. Accordingly, clinicians need to be aware that by doing so, breast cancer risk in families with DCIS diagnoses, will be probably overestimated.

Another issue that needs to be addressed, is that the PRS is widely validated in the European population, but not for all populations. We have validated the BOADICEA model including the PRS_{313} for the Dutch population. However, we selected for European ancestry while a substantial proportion of the Dutch population is of non-European ancestry. In the Netherlands at least 14% of the population was born themselves outside Western-Europe or one of their parents was born outside Western-Europe (Turkey or a country in Africa, South America or Asia)[25]. This means that we have validated the BOADICEA model for only ~86% of the Dutch population. The lack of ethnic diversity in genetic studies is a known problem. In example, of all included individuals in Genome Wide Association Studies (GWAS) until 2018, 78% are European, 10% are Asian, 2% are African, 1% are Hispanic, and all other ethnicities represent <1%[26]. Because of differences in linkage disequilibrium (LD) across ethnicities, it is uncertain if a causal variant is captured for all populations by the variant identified in GWAS of a single population. Related to this, it is known that some variants may be a risk factor in one population but protective in another population, a phenomenon termed flip-flop[27]. This phenomenon may be due to not targeting the true causal variant.

Table 1. Comparison of PRS performance for predicting overall breast cancer among different ancestries

Reference	PRS	Ancestry	Cases	Controls	OR per SD (95% CI)	AUC (95% CI)
Shieh et al. 2019[28]	180 variants	US Latinas and Latin American women	4,658	7,622	1.58 (1.52-1.64)	0.63 (0.62-0.64)
Ho et al. 2020[29]	287 variants[a]	European	11,225	17,788	1.61 (1.57–1.66)	0.63
		Asian	15,755	16,483	1.52 (1.49–1.56)	0.61
		Asians within North American	1,507	1,212	1.36 (1.25–1.49)	0.58
		Chinese	5,236	5,156	1.58 (1.51–1.65)	0.62 (0.60-0.63)
		Malay	1,084	1,332	1.48 (1.36–1.62)	0.60 (0.58-0.60)
		Indian	580	1,018	1.48 (1.33–1.65)	0.61 (0.59-0.64)
Du et al. 2021[30]	313 variants	African	9,241	10,193	1.27 (1.23-1.31)	0.57 (0.56-0.58)
Liu et al. 2021[31]	209 variants[a]	European	3,960	29,634	1.36 (1.31-1.41)	0.59 (0.58-0.60)
		African	274	3,527	1.15 (1.03-1.30)	0.53 (0.50-0.57)
		Latinx	147	2,049	1.20 (1.01-1.42)	0.53 (0.48-0.58)

[a]out of 313 variants as published by Mavaddat et al.[14]
Abbreviations: AUC, Area Under the Curve; CI, Confidence Interval; PRS, Polygenic Risk Score

Fortunately, there is growing attention for the underrepresentation of ethnically diverse populations in human genetics studies. In recent years, more work has been performed to determine the PRS performance in non-European ethnicities[28-32]. For the Asian population[29] and Latinas[28] the PRS showed similar performance as in the European population, but for the African population[30] there was clearly an attenuated effect size (Table 1). This latter may be due to the large heterogeneity in the African population leading to more variation in LD patterns across the continent[33]. Mapping of the true causal variants may help to obtain a more uniform PRS, useful for different ethnicities. Further research needs to be performed to make optimal use of the PRS for all individuals visiting clinical genetic services. Until more knowledge is gained about the performance of the PRS in women of other ethnicities or ethnicity-specific PRS are available, we have to be cautious when using comprehensive risk prediction including a European ancestry based PRS for these women.

7.3 Contralateral BC

7.3.1. Non-pathogenic variant carriers

In both the family studies described in **chapters 2 and 3** and the population-based studies in **chapters 2 and 4**, the PRS was on average higher for women who developed

a second primary breast tumour compared to women who developed a single breast tumour. These findings suggest an association of the PRS with the development of a second breast tumour which is indeed described in the literature for contralateral breast cancer[6, 34, 35]. However, the effect size of this association was weaker than found for a first breast cancer[35].

7.3.2. *BRCA1/2* pathogenic variant carriers

Previous research showed that the PRS was associated with breast cancer risk in women who carry a pathogenic variant in *BRCA1* or *BRCA2*[36, 37], although with a lower effect size compared to the population[14, 36]. Whether the PRS is associated with contralateral breast cancer risk for *BRCA1/2* pathogenic variant carriers as well, had not been investigated previously. In **chapter 5**, we investigated whether the PRS_{313} is associated with contralateral breast cancer risk among women of European ancestry who carry a pathogenic variant in *BRCA1* or *BRCA2* and explored the implications for contralateral breast cancer risk prediction for these women.

We used retrospective cohort data from carriers of a *BRCA1* or *BRCA2* pathogenic variant participating in the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA)[38] of which we included women of European ancestry with a prevalent first primary invasive breast cancer. We showed significant associations among both *BRCA1* and *BRCA2* pathogenic variant carriers between the PRS and contralateral breast cancer risk. However, as seen for the general population[35], the magnitude of the effect sizes were smaller than previously reported for the first breast cancer[36]. For *BRCA1* pathogenic variant carriers, the largest association was seen with the ER-negative PRS_{313} (HR per SD=1.12), while for *BRCA2* pathogenic variant carriers, both the PRS_{313} and ER-positive PRS_{313} showed similar associations with contralateral breast cancer risk (HR per SD=1.15). These findings are consistent with the higher relative prevalence in this cohort of ER-negative and ER-positive contralateral breast cancers respectively. Although the relative risks of the PRS for contralateral breast cancer were modest, differences in the PRS may still have an important effect on the absolute risk, which is high in BRCA1/2 pathogenic variant carriers. Therefore, the PRS could be used to refine estimates of contralateral breast cancer risks in women who carry a *BRCA1* or *BRCA2* pathogenic variant.

For both *BRCA1* and *BRCA2* pathogenic variant carriers, the strength of the association was greater for ER-positive contralateral breast cancers compared to ER-negative contralateral breast cancers, even if the ER-negative PRS was used for *BRCA1* pathogenic variant carriers. The effect sizes for the PRS are also larger for ER-positive disease in the general population, probably because ER-positive disease is commoner given that >75% of all breast tumours are ER-positive[39]. The same distribution holds for *BRCA2* pathogenic variant carriers as seen in our cohort and described in literature[40]. For *BRCA1* pathogenic variant carriers it is

the opposite, about 75-80% of the tumours are ER-negative[40]. In general, the effect size of the PRS_{313} for developing a first breast cancer[36] and contralateral breast cancer is similar for *BRCA1* and *BRCA2* pathogenic variant carriers. However, our results have shown that the PRS in carriers is mainly associated with ER-positive contralateral tumours and just slightly with ER-negative contralateral tumours. Given this, do we predict contralateral breast cancer risk well enough for *BRCA1* pathogenic variant carriers, or are we predicting only ER-positive contralateral breast cancer? For the first tumour, the ER-negative PRS showed good performance for predicting ER-negative tumours[36], therefore a pragmatic solution for *BRCA1* pathogenic variant carriers is to use the ER-negative PRS for risk prediction of the first tumour. However, this is not yet implemented in breast cancer risk prediction models such as BOADICEA[13]. Another solution would be to predict risks for ER-negative and ER-positive tumours separately. This could also inform clinical management, for example in guiding the choice for chemoprevention in case of an high risk for ER-positive tumour development[41, 42]. With larger datasets, it should be possible to develop better subtype specific PRS for breast cancer and contralateral breast cancer and use this PRS for clinical management choices.

Although the PRS may refine contralateral breast cancer risk estimates for women carrying a pathogenic variant in *BRCA1* or *BRCA2*, the effect size of the PRS seemed to decline with a higher age of first breast cancer diagnosis. For women who were diagnosed with a first tumour after the age of 50, the PRS was of less value for risk prediction for *BRCA2* pathogenic variant carriers and of no value for *BRCA1* pathogenic variant carriers. The decline of the effect size with higher age was seen as well for a first breast cancer for *BRCA1/2* pathogenic variant carriers[36] and for a first breast cancer in the general population[4, 14, 43] including our cohort described in **chapter 4.** However, there was some evidence that the decline in effect size was not linear, given the lower effect size below the age of 40 years described by Mavaddat et al.[14]. This effect was also seen in the population for a contralateral breast tumour[35]. The overall decline with higher age may be caused by a dilution of the effect size due to other risk factors, given that the risk for developing breast cancer in general increases with higher age. For age-dependent breast cancer risk prediction (i.e. 5-year risk or 10-year risk), it is important to take this age-effect into account.

7.4. Future perspectives

This thesis describes the clinical utility of using the PRS for individual breast cancer risk prediction. We have validated the association of the PRS with breast cancer for women in both the Dutch population and breast cancer families and showed a better risk-discrimination by adding the PRS to family-based risk prediction. Although the discrimination accuracy is modest with an AUC<0.70, it is an improvement compared to

family-based risk prediction. Secondly, we have shown that addition of the PRS to family-based risk prediction has an impact on screening recommendations for many non-carriers and carriers of a pathogenic variant in a moderate breast cancer gene. Lastly, there is a prospectively calibrated and externally validated model, BOADICEA, which gained approval as medical device (CE marking) and is implemented in the user-friendly web-interface, the CanRisk tool[44], to calculate breast cancer lifetime risks on the basis of genetic and non-genetic risk factors, including the PRS. The currently ongoing debate whether BOADICEA or other such models (e.g. Tyrer-Cuzick[45]) are good enough for implementation in the clinic and in the population screening setting will no doubt continue for some time; statistical modelling studies have suggested the efficacy of risk-based over age-based screening[46, 47], but these await real-life data from any of the several currently ongoing trials[48, 49] investigating the effect of risk-based screening in (semi-) randomised way.

In my opinion, we are ready for implementation of comprehensive risk prediction in clinical genetic services. However, exactly how to implement this new way of risk prediction has not yet materialised in detail. There remain many issues to be resolved and practicalities to be explored.

First, we have to explore if clinicians are ready to work with comprehensive risk prediction in their consultations. A recent study exploring the acceptability of the CanRisk tool, showed that it was generally acceptable to clinicians, but they were apprehensive about the impact of using this tool on their consultations, which can have impact on the level of implementation[50]. Clinicians are confident with screening advice recommendations based on family history-based risk prediction. As described in this thesis (chapter 2, 3, and 6), for a significant number of women, breast-cancer risks calculated including the PRS in CanRisk will be inconsistent with the risk category and corresponding clinical management advice based on family history only. Before implementation, clinicians must gain confidence in comprehensive risk prediction and corresponding results and we have to explore the effects amongst clinicians regarding their willingness to adjust current advises, especially when screening advices will be downgraded. This latter may also be important for the cost-effectiveness of comprehensive risk prediction. Related to this, we need to explore if comprehensive risk prediction will lead to differences in primary and/or secondary prevention choices by women. Furthermore, comprehensive risk prediction can result in a different screening advice for two family members, for example the two sisters shown in chapter 6. We have to explore the psychosocial effects of personal comprehensive risk prediction if the clinical management advice differs within a family in order to be able to anticipate on these effects. To conclude, before implementation of comprehensive risk prediction we need to know the acceptation of downgrading and different screening advice within families for both clinicians and counselees.

Secondly, due to bias towards European ancestry of Genome Wide Association Studies[26] as described above, the PRS is not yet validated for all ethnicities which may lead to health inequalities[51], resulting in an ethical challenge surrounding implementation of comprehensive risk prediction. Can we offer comprehensive risk prediction to women of European ancestry, if this is not yet possible for all women of non-European ancestries? Ideally, the same care is offered to all women in the population. However, because of other issues to be resolved before implementation, it is possible to start with a small group of women of European ancestry in research-setting to explore the ethical, psychosocial and logistical challenges of implementation. In the meantime, effort has to be made in human genetic research to validate the existing PRS_{313} in other ancestries or to determine ethnicity specific common low risk variants to compute ethnicity specific PRS. In the coming years, the Confluence project will address this by developing a large research resource to uncover breast cancer genetics through genome-wide association studies (GWAS) including cases and controls of different ethnicities[52]. It will be of added value if these results will be implemented in risk prediction models, for example by enabling inclusion of different effect sizes for the PRS.

Currently, if no pathogenic variant is detected in a family, the affected counselee will receive a family letter including the clinical advice for their healthy relatives. A practical issue for implementation of comprehensive risk prediction is that these unaffected relatives need to be referred for counselling for DNA sample collection and risk communication. In addition to the fact that comprehensive risk prediction is still time consuming, this will result in a lot more referrals to clinical genetic services for which we may not have the capacity at this moment. It would be helpful to invest in tools to speed up the process, for example by using pedigree data collection procedures that can be exported into the family tree structures that can be directly uploaded in the CanRisk tool.

Another practical issue is the development of a laboratory test to determine the PRS. This can be performed by direct genotyping each SNP separately, or by using a SNP array with additional imputation of the missing variants. Direct genotyping is technically easier, more efficient, and an advantage is the high reliability of the PRS calculation. Therefore, at this moment, laboratories prefer direct genotyping. However, in my opinion, using a genome-wide SNP-array and imputation has advantages that need to be seriously considered. A SNP array will be more future proof and widely applicable, given the possibility to calculate all kinds of PRS, not just those currently known for breast cancer and European ancestry. For example, it is to be expected that we will have a more extensive PRS for breast cancer in the future, knowing that the current PRS explains about half of the estimated part of the familial relative risk that could be explained by common low risk variants[14] and that recent studies already discovered 38 novel breast cancer susceptibility loci at genome wide significance level[53, 54]. Furthermore, although this is not yet implemented in the

CanRisk tool either, it is possible that we need to use ethnicity-specific PRS rather than adjustment of the weights associated with each variant of the PRS_{313}. In addition, because it may be difficult to define ancestry from non-genetic data (e.g. pedigree or anamnestic information), ancestry can be determined fairly accurately with array data. Finally, a sufficiently dense SNP array can also support the many other PRSs that have been defined today for many other common diseases, such as coronary arterial disease. In summary, direct genotyping might be favoured technically, but it is possible that we have to design multiple genetic test for all different PRS and need to estimate ancestry with non-genetic information.

While there are many challenges still to overcome, we can start in research-setting with implementation of individual comprehensive breast cancer risk prediction including the PRS for women visiting clinical genetic services and their healthy relatives. Hopefully the studies described in this thesis contribute to the first steps towards implementation of comprehensive risk prediction to all women in clinical setting and in the future for the population screening as well.

References

1. 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
2. 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]
3. 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
4. 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]
5. 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
6. 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]
7. 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
8. 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
9. 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
10. 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
11. Lee AJ, Cunningham AP, Tischkowitz M, et al. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. *Genetics in medicine : official journal of the American College of Medical Genetics*. Dec 2016;18(12):1190-1198. doi:10.1038/gim.2016.31
12. IKNL. Richtlijn Borstkanker - Screening buiten het bevolkingsonderzoek. Accessed 03-12-2021, https://richtlijnendatabase.nl/richtlijn/borstkanker/screening/screening_buiten_het_bob/screening_buiten_het_bevolkingsonderzoek.html

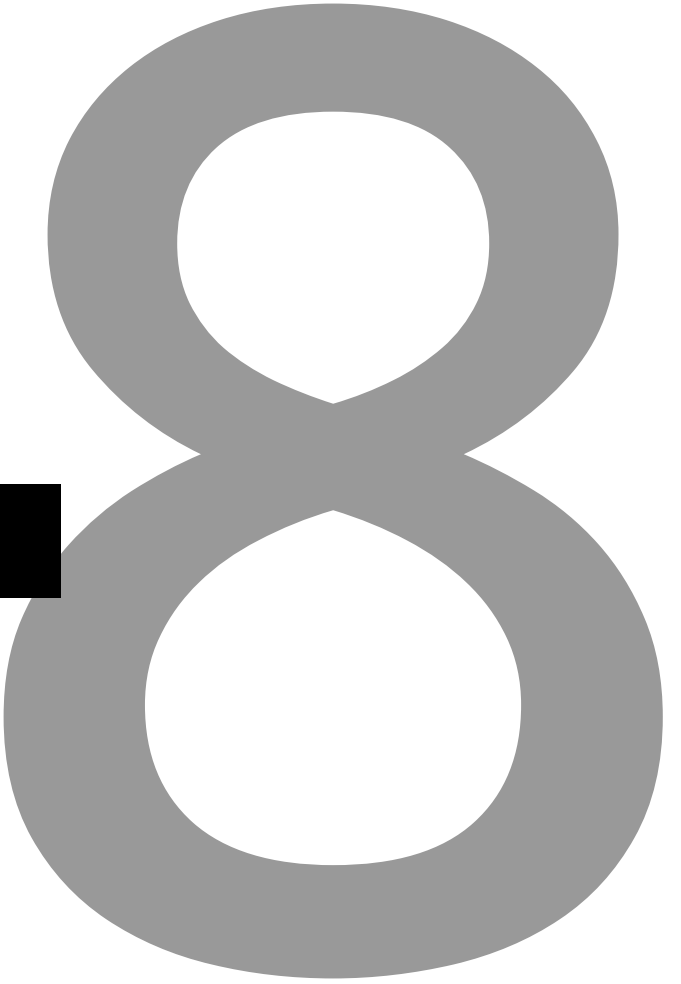
13. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in medicine : official journal of the American College of Medical Genetics*. Jan 15 2019;21(8):1708-1718. doi:10.1038/s41436-018-0406-9
14. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
15. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. *The Lancet Oncology*. Apr 2019;20(4):504-517. doi:10.1016/s1470-2045(18)30902-1
16. Pal Choudhury P, Brook MN, Hurson AN, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. *Breast Cancer Research*. 2021/02/15 2021;23(1):22. doi:10.1186/s13058-021-01399-7
17. Li SX, Milne RL, Nguyen-Dumont T, et al. Prospective Evaluation over 15 Years of Six Breast Cancer Risk Models. *Cancers*. 2021;13(20):5194.
18. Naslund-Koch C, Nordestgaard BG, Bojesen SE. Common breast cancer risk alleles and risk assessment: A study on 35,441 individuals from the Danish general population. *Annals of oncology : official journal of the European Society for Medical Oncology*. Oct 13 2016;doi:10.1093/annonc/mdw536
19. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA oncology*. Oct 1 2016;2(10):1295-1302. doi:10.1001/jamaoncol.2016.1025
20. Petridis C, Brook MN, Shah V, et al. Genetic predisposition to ductal carcinoma in situ of the breast. *Breast cancer research : BCR*. Feb 17 2016;18(1):22. doi:10.1186/s13058-016-0675-7
21. Visser LL, Groen EJ, van Leeuwen FE, Lips EH, Schmidt MK, Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. May 2019;28(5):835-845. doi:10.1158/1055-9965.Epi-18-0976
22. Groen EJ, Hudecek J, Mulder L, et al. Prognostic value of histopathological DCIS features in a large-scale international interrater reliability study. *Breast cancer research and treatment*. Oct 2020;183(3):759-770. doi:10.1007/s10549-020-05816-x
23. Coleman WB. Breast Ductal Carcinoma in Situ: Precursor to Invasive Breast Cancer. *Am J Pathol*. May 2019;189(5):942-945. doi:10.1016/j.ajpath.2019.03.002
24. van Seijen M, Lips EH, Thompson AM, et al. Ductal carcinoma in situ: to treat or not to treat, that is the question. *British journal of cancer*. Aug 2019;121(4):285-292. doi:10.1038/s41416-019-0478-6
25. CBS. Hoeveel mensen met een migratieachtergrond wonen in Nederland? Accessed 07-12-2021, https://www.cbs.nl/nl-nl/dossier/dossier-asiel-migratie-en-integratie/hoeveel-mensen-met-een-migratieachtergrond-wonen-in-nederland-#:~:text=Van%20de%20totale%20Nederlandse%20bevolking,daarmee%20tot%20de%20tweede%20generatie.

26. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell*. Mar 21 2019;177(1):26-31. doi:10.1016/j.cell.2019.02.048
27. Wang S, Qian F, Zheng Y, et al. Genetic variants demonstrating flip-flop phenomenon and breast cancer risk prediction among women of African ancestry. *Breast cancer research and treatment*. Apr 2018;168(3):703-712. doi:10.1007/s10549-017-4638-1
28. Shieh Y, Fejerman L, Lott PC, et al. A polygenic risk score for breast cancer in U.S. Latinas and Latin-American women. *Journal of the National Cancer Institute*. Sep 25 2019;doi:10.1093/jnci/djz174
29. Ho WK, Tan MM, Mavaddat N, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun*. Jul 31 2020;11(1):3833. doi:10.1038/s41467-020-17680-w
30. Du Z, Gao G, Adedokun B, et al. Evaluating Polygenic Risk Scores for Breast Cancer in Women of African Ancestry. *Journal of the National Cancer Institute*. Mar 26 2021;doi:10.1093/jnci/djab050
31. Liu C, Zeinomar N, Chung WK, et al. Generalizability of Polygenic Risk Scores for Breast Cancer Among Women With European, African, and Latinx Ancestry. *JAMA Netw Open*. Aug 2 2021;4(8):e2119084. doi:10.1001/jamanetworkopen.2021.19084
32. Evans DG, van Veen EM, Byers H, et al. The importance of ethnicity: Are breast cancer polygenic risk scores ready for women who are not of White European origin? *International journal of cancer*. Jan 1 2022;150(1):73-79. doi:10.1002/ijc.33782
33. Tishkoff SA, Reed FA, Friedlaender FR, et al. The genetic structure and history of Africans and African Americans. *Science (New York, NY)*. May 22 2009;324(5930):1035-44. doi:10.1126/science.1172257
34. 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
35. Kramer I, Hooning MJ, Mavaddat N, et al. Breast Cancer Polygenic Risk Score and Contralateral Breast Cancer Risk. *American journal of human genetics*. Nov 5 2020;107(5):837-848. doi:10.1016/j.ajhg.2020.09.001
36. Barnes DR, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in medicine : official journal of the American College of Medical Genetics*. Oct 2020;22(10):1653-1666. doi:10.1038/s41436-020-0862-x
37. 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
38. Chenevix-Trench G, Milne RL, Antoniou AC, Couch FJ, Easton DF, Goldgar DE. An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). *Breast cancer research : BCR*. 2007;9(2):104. doi:10.1186/bcr1670

39. Tsang JYS, Tse GM. Molecular Classification of Breast Cancer. *Adv Anat Pathol*. Jan 2020;27(1):27-35. doi:10.1097/pap.0000000000000232
40. Spurdle AB, Couch FJ, Parsons MT, et al. Refined histopathological predictors of BRCA1 and BRCA2 mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. *Breast cancer research : BCR*. Dec 23 2014;16(6):3419. doi:10.1186/s13058-014-0474-y
41. Nelson HD, Fu R, Griffin JC, Nygren P, Smith ME, Humphrey L. Systematic review: comparative effectiveness of medications to reduce risk for primary breast cancer. *Annals of internal medicine*. Nov 17 2009;151(10):703-15, w-226-35. doi:10.7326/0003-4819-151-10-200911170-00147
42. Li K, Anderson G, Viallon V, et al. Risk prediction for estrogen receptor-specific breast cancers in two large prospective cohorts. *Breast cancer research : BCR*. Dec 3 2018;20(1):147. doi:10.1186/s13058-018-1073-0
43. Gao C, Polley EC, Hart SN, et al. Risk of Breast Cancer Among Carriers of Pathogenic Variants in Breast Cancer Predisposition Genes Varies by Polygenic Risk Score. *Journal of Clinical Oncology*. 0(0):JCO.20.01992. doi:10.1200/jco.20.01992
44. Carver T, Hartley S, Lee A, et al. CanRisk Tool—A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer Epidemiology Biomarkers & Prevention*. 2021;30(3):469-473. doi:10.1158/1055-9965.Epi-20-1319
45. Kim G, Bahl M. Assessing Risk of Breast Cancer: A Review of Risk Prediction Models. *J Breast Imaging*. Mar-Apr 2021;3(2):144-155. doi:10.1093/jbi/wbab001
46. van den Broek JJ, Schechter CB, van Ravesteyn NT, et al. Personalizing Breast Cancer Screening Based on Polygenic Risk and Family History. *Journal of the National Cancer Institute*. Apr 6 2021;113(4):434-442. doi:10.1093/jnci/djaa127
47. Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. *JAMA oncology*. Nov 1 2018;4(11):1504-1510. doi:10.1001/jamaoncol.2018.1901
48. Clift AK, Dodwell D, Lord S, et al. The current status of risk-stratified breast screening. *British journal of cancer*. Oct 26 2021;doi:10.1038/s41416-021-01550-3
49. Pashayan N, Antoniou AC, Ivanus U, et al. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol*. Nov 2020;17(11):687-705. doi:10.1038/s41571-020-0388-9
50. Archer S, Babb de Villiers C, Scheibl F, et al. Evaluating clinician acceptability of the prototype CanRisk tool for predicting risk of breast and ovarian cancer: A multi-methods study. *PloS one*. 2020;15(3):e0229999. doi:10.1371/journal.pone.0229999
51. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*. Apr 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x
52. Confluence project. Accessed 18-01-2022, 2022. https://dceg.cancer.gov/research/cancer-types/breast-cancer/confluence-project

53. Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature genetics*. Jun 2020;52(6):572-581. doi:10.1038/s41588-020-0609-2
54. Adedokun B, Du Z, Gao G, et al. Cross-ancestry GWAS meta-analysis identifies six breast cancer loci in African and European ancestry women. *Nat Commun*. Jul 7 2021;12(1):4198. doi:10.1038/s41467-021-24327-x

CHAPTER 8



Nederlandse samenvatting
List of publications
Curriculum Vitae
Dankwoord

Nederlandse samenvatting

Achtergrond

Borstkanker is wereldwijd de meest voorkomende kanker bij vrouwen[1]. In Nederland hebben vrouwen gemiddeld een risico van 13% om gedurende het leven borstkanker te ontwikkelen[2]. De hoge incidentie zorgde in 1990 voor de start van het bevolkingsonderzoek, waarbij vrouwen boven de 50 jaar elke twee jaar uitgenodigd worden voor screening middels een mammografie. Sinds de invoering van het bevolkingsonderzoek is het sterftecijfer voor borstkanker gedaald, mogelijk mede door verbetering van de behandelmogelijkheden[3, 4]. Daartegenover staat het twee keer zoveel vinden van borstkanker[5], deels tumoren die anders nooit klinisch relevant waren geworden (over-diagnose)[6, 7]. Daarnaast blijkt een mammogram soms fout-positief te zijn na het verrichten van een biopsie[8, 9]. Op dit moment is de keuze voor de start van screening voor vrouwen alleen op leeftijd gebaseerd. Vanaf 50 jaar is het risico hoog genoeg om screening kosteneffectief te maken[10]. Door rekening te houden met andere relevante risicofactoren naast leeftijd, wordt er mogelijk een meer optimale verhouding verkregen tussen de voor- en nadelen van screening[11].

Een van de belangrijkste risicofactoren voor borstkanker is het hebben van een positieve familiegeschiedenis voor borstkanker[12]. Een vrouw die minstens één eerstegraads familielid heeft met borstkanker, heeft zelf ongeveer een twee keer zo hoog risico om borstkanker te ontwikkelen ten opzichte van een vrouw die dit niet heeft. Dit relatieve risico van ~2 wordt het familiair relatief risico genoemd en wordt deels verklaard (~25%) door pathogene (ziekte-veroorzakende) varianten in borstkankergenen (*BRCA1/2*, *PALB2*, *ATM* en *CHEK2*). Dragers van pathogene varianten in het *BRCA1/2* of *PALB2* gen hebben een hoog risico (45-80%) en dragers van pathogene varianten in het *ATM* of *CHEK2* gen hebben een matig verhoogd risico (20-45%) om gedurende het leven borstkanker te ontwikkelen.

Een ander belangrijk deel van het familiair relatief risico (~18%) wordt verklaard door veel voorkomende varianten (risico-allelen) die geassocieerd zijn met een laag risico voor borstkanker[13, 14]. De verwachting is dat een nog groter percentage, in totaal ~40%, van het risico verklaard zou kunnen worden door deze risico-allelen[14]. Naar schatting gaat het om enkele honderden, misschien wel duizenden van dergelijke allelen. Dit hoge aantal veel voorkomende varianten en de normaal verdeling in de populatie, maakt dat alle personen drager zijn van een bepaalde hoeveelheid van deze varianten, in tegenstelling tot pathogene varianten in de bovengenoemde borstkanker genen. De meeste mensen in de populatie zijn drager van een gemiddeld aantal risico-allelen waardoor het risico op borstkanker gelijk zal zijn aan het gemiddeld risico in de populatie, namelijk ~13%. Een deel van de mensen in de populatie zal echter drager zijn van minder risico-allelen

en een deel van meer risico-allelen, leidend tot een lager en hoger risico respectievelijk in vergelijking met het populatie risico. Individueel geven deze risico-allelen maar een kleine toename van het risico op borstkanker, maar hun gezamenlijke effect kan aanzienlijk hoger zijn[15]. De zogenaamde Polygene Risico Score (PRS) is een samenvattende risico score van al deze allelen en het allel-specifieke risico samen. In eerdere studies is gebleken dat de PRS bruikbaar kan zijn voor het verdelen van vrouwen in verschillende risicocategorieën[14-19]. Dit biedt de mogelijkheid om het risico en screeningsadvies te personaliseren voor vrouwen uit borstkankerfamilies. Daarnaast kan de PRS nuttig zijn bij het verfijnen van het risico voor vrouwen die drager zijn van een pathogene variant in een van de borstkankergenen[20-23].

In de huidige klinisch genetische praktijk wordt nog geen gebruik gemaakt van de PRS. Bij vrouwen die op jonge leeftijd borstkanker hebben ontwikkeld of meerdere familieleden hebben met borstkanker, wordt DNA onderzoek verricht naar tenminste de vijf borstkankergenen *BRCA1/2*, *PALB2*, *ATM* en *CHEK2*. In het merendeel van de borstkankerfamilies wordt geen pathogene variant aangetoond. Om wel een passend screeningsadvies te geven aan gezonde vrouwen in deze families, wordt het borstkankerrisico bepaald op basis van de aangedane en gezonde vrouwen in de familie. Hierbij kan gebruik gemaakt worden van verschillende risicopredictie modellen[24], zoals de Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA)[25]. In de huidige Nederlandse richtlijn bestaan drie risicocategorieën waarop het screeningsadvies bepaald wordt (laag, gemiddeld en hoog risico (hoofdstuk 1, tabel 2)).

Doel van het onderzoek

Het belangrijkste doel van het onderzoek beschreven in dit proefschrift is het onderzoeken van de klinische bruikbaarheid van de PRS voor individuele voorspelling van het risico op borstkanker. Dit hebben we gedaan door kennis te genereren over de PRS in de Nederlandse algemene bevolking, in borstkankerfamilies en in een groot internationaal cohort van vrouwelijke dragers van een pathogene variant in het *BRCA1* of *BRCA2* gen. Met de resultaten uit dit onderzoek hopen we de implementatie van de PRS in de kliniek te ondersteunen zodat vrouwen beter geïnformeerde keuzes kunnen maken over de mogelijkheden voor borstkanker preventie.

Resultaten

In **hoofdstukken 2 en 3** onderzochten we de klinische toepasbaarheid van de PRS voor risicovoorspelling voor vrouwen uit borstkankerfamilies. De vrouwen geïncludeerd in deze studies zijn bij een van de klinisch genetische centra in Nederland geweest voor counseling waarbij met DNA-diagnostiek geen pathogene variant in *BRCA1* of *BRCA2* is gevonden.

In **hoofdstuk 2** zijn zowel aangedane als gezonde vrouwen uit 101 sterk belaste families geïncludeerd. De PRS in deze studie was gebaseerd op 161 borstkanker risico-allelen. Binnen dit cohort hadden vrouwen met borstkanker gemiddeld een hogere PRS in vergelijking met hun gezonde vrouwelijke familieleden. Binnen deze families hebben we een associatie bevestigd tussen het voorkomen van borstkanker en een hogere PRS. Het toevoegen van de PRS aan het borstkankerrisico op basis van de familiegeschiedenis, wat in de huidige klinische praktijk gebruikt wordt, zorgde in 20% van zowel de aangedane als de niet-aangedane vrouwen voor verandering van risicocategorie. Met deze verandering zouden ze een ander screeningsadvies hebben gekregen op basis van de Nederlandse richtlijn borstkankerscreening[26].

In **hoofdstuk 3** zijn 3,918 vrouwen met borstkanker geïncludeerd uit 3,501 borstkanker families. Deze groep was meer representatief voor de borstkankerfamilies die in de huidige klinisch genetische praktijk gezien worden. Daarnaast werd deze groep nu vergeleken met gezonde controles uit de algemene populatie in plaats van gezonde familieleden. De PRS werd berekend op basis van 313 risico-allelen (PRS_{313}). Met de informatie uit de stambomen (aangedane en gezonde vrouwen), werd met het BOADICEA model een score berekend op basis van de familiegeschiedenis. Zoals ook aangetoond in **hoofdstuk 2** en beschreven in de literatuur[19, 27], was er slechts een zeer zwakke positieve correlatie tussen de PRS en deze score. In families waar meer dan één familielid was geïncludeerd, kon 13% van de variantie in de PRS_{313} worden verklaard door de PRS_{313} van diegene met de jongste borstkanker diagnose. Deze resultaten benadrukken de toegevoegde waarde van het bepalen van de PRS voor elk individu in de familie: de PRS kan immers nauwelijks afgeleid of geschat worden aan de hand van de informatie over de familiegeschiedenis. Net als in **hoofdstuk 2**, hebben we ook bij deze groep vrouwen met borstkanker uit borstkankerfamilies de associatie bevestigd tussen het voorkomen van borstkanker en een hogere PRS. De associatie was hoger voor invasief borstkanker ten opzichte van in situ borstkanker (voorstadium). Voor de meerderheid werd DNA onderzoek verricht naar de genen *CHEK2*, *ATM* en *PALB2*. In totaal was 1,8% van de controles en 8,4% van de aangedane vrouwen drager van een pathogene variant in een van deze genen, het meest frequent in *CHEK2*. Met behulp van BOADICEA versie 5, waarin de PRS_{313} kan worden toegevoegd[25], werden borstkankerrisico's berekend met en zonder de PRS_{313}. Het toevoegen van de PRS_{313} aan het borstkankerrisico op basis van de familiegeschiedenis en uitslag van het genen panel zorgde voor een verschuiving van risicocategorie voor maximaal 34% van de aangedane vrouwen zonder pathogene variant in een van de genen. Het had eveneens een grote impact op de verschuiving van risicocategorie voor dragers van een *ATM*- en *CHEK2* pathogene variant, wat overeenkomt met het gesuggereerde polygene effect van deze genen. Er werd geen verandering gevonden voor dragers van een *PALB2* pathogene variant, die allemaal in de hoog risico categorie bleven. Variaties in risicoscores voor

deze dragers zouden echter wel van invloed kunnen zijn op keuzes met betrekking tot profylactische chirurgie.

In **hoofdstuk 4** werd gebruik gemaakt van een groot bestaand prospectief cohort van vrouwen uit de Nederlandse populatie. Deze vrouwen waren 45 jaar of ouder met uitgebreide follow-up gegevens tot 25 jaar na inclusie. In deze studie onderzochten we de bruikbaarheid van BOADICEA en de associatie van borstkanker met de PRS_{313} voor de algemene Nederlandse bevolking. Vrouwen die borstkanker ontwikkelden, hadden gemiddeld een hogere PRS_{313} in vergelijking met niet-aangedane vrouwen. Verder, zoals eerder gevonden in **hoofdstuk 3**, hadden vrouwen die een invasieve borsttumor ontwikkelden gemiddeld een hogere PRS_{313} dan vrouwen met een in situ borsttumor. Een hogere PRS_{313} was geassocieerd met het ontwikkelen van borstkanker, met een vergelijkbare effectgrootte als in een eerdere prospectieve studie van Europese vrouwen[14]. Dit resultaat toont de robuustheid van het effect van de PRS en potentiële toepassing op de Nederlandse bevolking aan. De PRS_{313} bleek tevens specifiek geassocieerd te zijn met borstkanker en niet met andere tumoren. Met behulp van het BOADICEA model werden cumulatieve 10-jaars borstkankerrisico scores berekend met behulp van leeftijd, risicofactoren en de PRS_{313}. Het onderscheidt tussen aangedane en niet-aangedane vrouwen was het beste te maken met behulp van de PRS zoals ook in eerder onderzoek is aangetoond[25, 28]. Het kon slechts marginaal verder worden verbeterd door het toevoegen van risicofactoren (levensstijl, reproductieve factoren en antropometrische gegevens). BOADICEA onderschatte wel het waargenomen 10-jaars borstkankerrisico van 4,4% in de totale groep vrouwen, vooral in de hoogste risicocategorieën. Deze onderschatting was mogelijk te wijten aan het ontbreken van gegevens over familiegeschiedenis, mammadensiteit en informatie over pathogene varianten in BRCA1/2. Over het algemeen lijkt het onderscheidende vermogen van het BOADICEA model geschikt voor implementatie in preventieprogramma's voor borstkanker, maar voor nauwkeurig gebruik kan het belangrijk zijn om informatie over de familiegeschiedenis en dragerschap van pathogene varianten in borstkankergenen toe te voegen.

In zowel de familiestudies beschreven in de **hoofdstukken 2 en 3** als de populatiestudies in de **hoofdstukken 2 en 4** was de PRS gemiddeld hoger voor vrouwen die een tweede primaire borsttumor ontwikkelden in vergelijking met vrouwen die een enkele borsttumor ontwikkelden. Deze bevindingen suggereren een associatie van een hogere PRS met de ontwikkeling van een tweede borsttumor zoals ook in de literatuur is beschreven[17, 29, 30].

Eerder onderzoek heeft aangetoond dat een hogere PRS ook geassocieerd is met het ontwikkelen van borstkanker bij vrouwen die drager zijn van een pathogene variant in het *BRCA1* of *BRCA2* gen[20, 31]. De grootte van deze associatie was wel lager in vergelijking met de associatie in de populatie[14, 31]. Of de PRS ook geassocieerd is met het ontwikkelen van

een tweede primaire tumor in de andere borst (contralateraal) voor *BRCA1/2* pathogene variant dragers was niet eerder onderzocht. In **hoofdstuk 5** hebben we Europese vrouwelijke dragers van een pathogene variant in het *BRCA1* of *BRCA2* gen geïncludeerd uit het Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) cohort[32]. De geïncludeerde vrouwen in onze studie hadden eerder een invasieve borsttumor ontwikkeld. Voor zowel dragers van een pathogene variant in het *BRCA1* als *BRCA2* gen werd een associatie aangetoond tussen het voorkomen van contralateraal borstkanker en een hogere PRS_{313}. Echter, zoals ook gezien voor vrouwen uit de algemene populatie[30], was de associatie kleiner dan eerder gerapporteerd voor de eerste borsttumor[31]. Hoewel de associatie relatief bescheiden was, kunnen verschillen in de PRS_{313} nog steeds een belangrijke invloed hebben op het absolute risico, dat hoog is bij dragers van pathogene BRCA1/2-varianten. Daarom zou de PRS_{313} gebruikt kunnen worden om het risico op contralateraal borstkanker bij deze vrouwen te verfijnen.

In **hoofdstuk 6** hebben we de resultaten van een pilot studie beschreven waarin we het risico op borstkanker hebben berekend en gecommuniceerd via een web consult aan 38 gezonde vrouwen. Deze vrouwen zijn eerstegraads familieleden van vrouwen met borstkanker die bekend zijn bij de klinische genetica en bij wie eerder geen pathogene variant in een van de borstkankergenen werd aangetoond. Het doel was om de klinische en emotionele impact te bepalen van het gebruik en de communicatie van uitgebreide risicovoorspelling (CRP, comprehensive risk prediction) met behulp van de familiegeschiedenis, PRS en risicofactoren (leefstijl-/hormonale factoren). Om de psychosociale impact te beoordelen, werd de deelnemers gevraagd vragenlijsten in te vullen voor en na risicocommunicatie. Bijna de helft van de vrouwen (47%) verschoof naar een andere risicocategorie en kreeg op basis van hun CRP een ander screeningadvies vergeleken met het eerder gegeven advies op basis van de familiegeschiedenis alleen. De deelnemers waren over het algemeen positief over het ontvangen van hun CRP, de uitleg daarbij en de manier van communiceren, namelijk online via een web consult.

Conclusie

In het onderzoek beschreven in dit proefschrift hebben we de associatie van de PRS met borstkanker voor vrouwen in zowel de Nederlandse algemene bevolking als in borstkankerfamilies gevalideerd. Toevoeging van de PRS maakt een beter onderscheid mogelijk tussen vrouwen met en zonder borstkanker. Hoewel de nauwkeurigheid van dit onderscheid nog bescheiden is, lijkt dit wel een verbetering ten opzichte van de huidige risicovoorspelling. Verder heeft het toevoegen van de PRS aan familiegeschiedenis gebaseerde risicoschatting een grote impact op screeningsadviezen voor zowel niet-dragers en dragers van een pathogene variant in het *ATM* of *CHEK2* gen. Ten slotte is er een gevalideerd risicopredictie model beschikbaar, BOADICEA, welke is geïmplementeerd in de gebruiksvriendelijke CanRisk-tool (www.canrisk.org)[33], waarmee het risico op

borstkanker berekend kan worden op basis van zowel genetische en niet-genetische risicofactoren. Deze resultaten suggereren dat we klaar zijn voor implementatie van de PRS. Echter, voordat implementatie mogelijk is zijn er nog veel (logistieke) uitdagingen aan te gaan zoals bepalen of het effect van de PRS hetzelfde is voor vrouwen van niet-Europese afkomst en het onderzoeken van de psychosociale effecten van een uitgebreide individuele risicovoorspelling indien zussen bijvoorbeeld een ander screeningsadvies zouden krijgen. Hopelijk dragen de studies beschreven in dit proefschrift bij aan de implementatie van uitgebreide risicovoorspelling voor borstkanker: zowel voor vrouwen uit borstkanker families, als voor vrouwen uit de algemene populatie.

Referenties

1. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European journal of cancer (Oxford, England : 1990)*. Nov 2018;103:356-387. doi:10.1016/j.ejca.2018.07.005
2. Kennisinstituut van Federatie Medisch Specialisten. Richtlijn Borstkanker. . Accessed October 13, 2021. https://richtlijnendatabase.nl/richtlijn/borstkanker/algemeen.html
3. Otto SJ, Fracheboud J, Looman CW, et al. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. *Lancet (London, England)*. Apr 26 2003;361(9367):1411-7. doi:10.1016/s0140-6736(03)13132-7
4. Tabár L, Dean PB, Chen TH, et al. The incidence of fatal breast cancer measures the increased effectiveness of therapy in women participating in mammography screening. *Cancer*. Feb 15 2019;125(4):515-523. doi:10.1002/cncr.31840
5. IKNL. Cijfers over kanker, Nederlandse kankerregistratie. Accessed October 12, 2021. http://www.cijfersoverkanker.nl
6. Ripping TM, Verbeek AL, Fracheboud J, de Koning HJ, van Ravesteyn NT, Broeders MJ. Overdiagnosis by mammographic screening for breast cancer studied in birth cohorts in The Netherlands. *International journal of cancer*. Aug 15 2015;137(4):921-9. doi:10.1002/ijc.29452
7. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *New England Journal of Medicine*. 2016;375(15):1438-1447. doi:10.1056/NEJMoa1600249
8. Bleyer A. Screening mammography: update and review of publications since our report in the New England Journal of Medicine on the magnitude of the problem in the United States. *Academic radiology*. Aug 2015;22(8):949-60. doi:10.1016/j.acra.2015.03.003
9. Myers ER, Moorman P, Gierisch JM, et al. Benefits and Harms of Breast Cancer Screening: A Systematic Review. *Jama*. Oct 20 2015;314(15):1615-34. doi:10.1001/jama.2015.13183
10. van der Waal D, Verbeek AL, den Heeten GJ, Ripping TM, Tjan-Heijnen VC, Broeders MJ. Breast cancer diagnosis and death in the Netherlands: a changing burden. *European journal of public health*. Apr 2015;25(2):320-4. doi:10.1093/eurpub/cku088
11. Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. *JAMA oncology*. Nov 1 2018;4(11):1504-1510. doi:10.1001/jamaoncol.2018.1901
12. 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
13. 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
14. 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*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002

15. 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]
16. Lakeman IMM, Hilbers FS, Rodriguez-Girondo M, et al. 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. *Journal of medical genetics*. Sep 2019;56(9):581-589. doi:10.1136/jmedgenet-2019-106072
17. 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]
18. 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
19. 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
20. 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
21. Barnes D, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in Medicine*. 2020;accepted for publication
22. Gao C, Polley EC, Hart SN, et al. Risk of Breast Cancer Among Carriers of Pathogenic Variants in Breast Cancer Predisposition Genes Varies by Polygenic Risk Score. *Journal of Clinical Oncology*. 0(0):JCO.20.01992. doi:10.1200/jco.20.01992
23. Gallagher S, Hughes E, Wagner S, et al. Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. *JAMA Network Open*. 2020;3(7):e208501-e208501. doi:10.1001/jamanetworkopen.2020.8501
24. 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
25. 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
26. IKNL. Richtlijn Borstkanker - Screening buiten het bevolkingsonderzoek. Accessed 03-12-2021, https://richtlijnendatabase.nl/richtlijn/borstkanker/screening/screening_buiten_het_bob/screening_buiten_het_bevolkingsonderzoek.html

27. 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]
28. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA oncology*. Oct 1 2016;2(10):1295-1302. doi:10.1001/jamaoncol.2016.1025
29. 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
30. Kramer I, Hooning MJ, Mavaddat N, et al. Breast Cancer Polygenic Risk Score and Contralateral Breast Cancer Risk. *American journal of human genetics*. Nov 5 2020;107(5):837-848. doi:10.1016/j.ajhg.2020.09.001
31. Barnes DR, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in medicine : official journal of the American College of Medical Genetics*. Oct 2020;22(10):1653-1666. doi:10.1038/s41436-020-0862-x
32. Chenevix-Trench G, Milne RL, Antoniou AC, Couch FJ, Easton DF, Goldgar DE. An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). *Breast cancer research : BCR*. 2007;9(2):104. doi:10.1186/bcr1670
33. Carver T, Hartley S, Lee A, et al. CanRisk Tool—A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer Epidemiology Biomarkers & Prevention*. 2021;30(3):469-473. doi:10.1158/1055-9965.Epi-20-1319

List of publications

Lakeman IMM, Hilbers FS, Rodriguez-Girondo M, Lee A, Vreeswijk MPG, Hollestelle A, Seynaeve C, Meijers-Heijboer H, Oosterwijk JC, Hoogerbrugge N, Olah E, Vasen HFA, van Asperen CJ, Devilee P. 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. Journal of medical genetics 2019;56(9):581-89. doi: 10.1136/jmedgenet-2019-106072

Lakeman IMM, Schmidt MK, van Asperen CJ, Devilee P. Breast Cancer Susceptibility—Towards Individualised Risk Prediction. Current Genetic Medicine Reports 2019;7(2):124-35. doi: 10.1007/s40142-019-00168-5

Lakeman IMM, Rodríguez-Girondo M, Lee A, Ruiter R, Stricker BH, Wijnant SRA, Kavousi M, Antoniou AC, Schmidt MK, Uitterlinden AG, van Rooij J, Devilee P. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. Genetics in medicine: official journal of the American College of Medical Genetics 2020;22(11):1803-11. doi: 10.1038/s41436-020-0884-4

Potjer TP, van der Grinten TWJ, **Lakeman IMM**, Bollen SH, Rodríguez-Girondo M, Iles MM, Barrett JH, Kiemeney LA, Gruis NA, van Asperen CJ, van der Stoep N. Association between a 46-SNP Polygenic Risk Score and melanoma risk in Dutch patients with familial melanoma. Journal of medical genetics 2020. doi: 10.1136/jmedgenet-2020-107251

Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, Pooley KA, Parsons MT, Fortuno C, Wang Q, Bolla MK, Dennis J, Keeman R, Alonso MR, Álvarez N, Herraez B, Fernandez V, Núñez-Torres R, Osorio A, Valcich J, Li M, Törngren T, Harrington PA, Baynes C, Conroy DM, Decker B, Fachal L, Mavaddat N, Ahearn T, Aittomäki K, Antonenkova NN, Arnold N, Arveux P, Ausems M, Auvinen P, Becher H, Beckmann MW, Behrens S, Bermisheva M, Białkowska K, Blomqvist C, Bogdanova NV, Bogdanova-Markov N, Bojesen SE, Bonanni B, Børresen-Dale AL, Brauch H, Bremer M, Briceno I, Brüning T, Burwinkel B, Cameron DA, Camp NJ, Campbell A, Carracedo A, Castelao JE, Cessna MH, Chanock SJ, Christiansen H, Collée JM, Cordina-Duverger E, Cornelissen S, Czene K, Dörk T, Ekici AB, Engel C, Eriksson M, Fasching PA, Figueroa J, Flyger H, Försti A, Gabrielson M, Gago-Dominguez M, Georgoulias V, Gil F, Giles GG, Glendon G, Garcia EBG, Alnæs GIG, Guénel P, Hadjisavvas A, Haeberle L, Hahnen E, Hall P, Hamann U, Harkness EF, Hartikainen JM, Hartman M, He W, Heemskerk-Gerritsen BAM, Hillemanns P, Hogervorst FBL, Hollestelle A, Ho WK, Hooning MJ, Howell A, Humphreys K, Idris F, Jakubowska A, Jung A, Kapoor PM, Kerin MJ, Khusnutdinova E, Kim SW, Ko YD, Kosma VM, Kristensen VN, Kyriacou K, **Lakeman IMM**, Lee JW, Lee MH, Li J, Lindblom A, Lo WY, Loizidou MA, Lophatananon A, Lubiński J, MacInnis RJ, Madsen MJ, Mannermaa A, Manoochehri M, Manoukian S, Margolin S, Martinez ME, Maurer T,

Mavroudis D, McLean C, Meindl A, Mensenkamp AR, Michailidou K, Miller N, Mohd Taib NA, Muir K, Mulligan AM, Nevanlinna H, Newman WG, Nordestgaard BG, Ng PS, Oosterwijk JC, Park SK, Park-Simon TW, Perez JIA, Peterlongo P, Porteous DJ, Prajzendanc K, Prokofyeva D, Radice P, Rashid MU, Rhenius V, Rookus MA, Rüdiger T, Saloustros E, Sawyer EJ, Schmutzler RK, Schneeweiss A, Schürmann P, Shah M, Sohn C, Southey MC, Surowy H, Suvanto M, Thanasitthichai S, Tomlinson I, Torres D, Truong T, Tzardi M, Valova Y, van Asperen CJ, Van Dam RM, van den Ouweland AMW, van der Kolk LE, van Veen EM, Wendt C, Williams JA, Yang XR, Yoon SY, Zamora MP, Evans DG, de la Hoya M, Simard J, Antoniou AC, Borg Å, Andrulis IL, Chang-Claude J, García-Closas M, Chenevix-Trench G, Milne RL, Pharoah PDP, Schmidt MK, Spurdle AB, Vreeswijk MPG, Benitez J, Dunning AM, Kvist A, Teo SH, Devilee P, Easton DF. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. The New England journal of medicine 2021. doi: 10.1056/NEJMoa1913948

Lakeman IMM, van den Broek AJ, Vos JAM, Barnes DR, Adlard J, Andrulis IL, Arason A, Arnold N, Arun BK, Balmaña J, Barrowdale D, Benitez J, Borg A, Caldés T, Caligo MA, Chung WK, Claes KBM, Collée JM, Couch FJ, Daly MB, Dennis J, Dhawan M, Domchek SM, Eeles R, Engel C, Evans DG, Feliubadaló L, Foretova L, Friedman E, Frost D, Ganz PA, Garber J, Gayther SA, Gerdes AM, Godwin AK, Goldgar DE, Hahnen E, Hake CR, Hamann U, Hogervorst FBL, Hooning MJ, Hopper JL, Hulick PJ, Imyanitov EN, Isaacs C, Izatt L, Jakubowska A, James PA, Janavicius R, Jensen UB, Jiao Y, John EM, Joseph V, Karlan BY, Kets CM, Konstantopoulou I, Kwong A, Legrand C, Leslie G, Lesueur F, Loud JT, Lubiński J, Manoukian S, McGuffog L, Miller A, Gomes DM, Montagna M, Mouret-Fourme E, Nathanson KL, Neuhausen SL, Nevanlinna H, Yie JNY, Olah E, Olopade OI, Park SK, Parsons MT, Peterlongo P, Piedmonte M, Radice P, Rantala J, Rennert G, Risch HA, Schmutzler RK, Sharma P, Simard J, Singer CF, Stadler Z, Stoppa-Lyonnet D, Sutter C, Tan YY, Teixeira MR, Teo SH, Teulé A, Thomassen M, Thull DL, Tischkowitz M, Toland AE, Tung N, van Rensburg EJ, Vega A, Wappenschmidt B, Devilee P, van Asperen CJ, Bernstein JL, Offit K, Easton DF, Rookus MA, Chenevix-Trench G, Antoniou AC, Robson M, Schmidt MK. The predictive ability of the 313 variant-based polygenic risk score for contralateral breast cancer risk prediction in women of European ancestry with a heterozygous BRCA1 or BRCA2 pathogenic variant. Genetics in medicine : official journal of the American College of Medical Genetics 2021;23(9):1726-37. doi: 10.1038/s41436-021-01198-7

Lakeman IMM, Rodriguez-Girondo M, Lee A, Celosse N, Braspenning M, Engelen Kv, Beek Ivd, Hout Avd, Garcia EG, Mensenkamp A, Ausems M, Hooning M, Adank MA, Hollestelle A, Schmidt M, Asperen Cv, Devilee P. Clinical applicability of the Polygenic Risk Score for breast cancer risk prediction in familial cases. *Submitted*

Brédart A, de Pauw A, Tüchler A, **Lakeman IMM**, Anota A, Rhiem K, Schmutzler R, van Asperen CJ, Devilee P, Stoppa-Lyonnet D, Kop J, Dolbeault S. Genetic clinicians' confidence

in counselling with BOADICEA comprehensive breast cancer risk estimates and counselees' psychosocial outcomes: an observational prospective study. *Submitted*

Dorling L, Carvalho S, Allen J, Parsons MT, Fortuno C, González-Neira A, Heijl SM, Adank MA, Ahearn TU, Andrulis IL, Auvinen P, Becher H, Beckmann MW, Behrens S, Bermisheva M, Bogdanova NV, Bojesen SE, Bolla MK, Bremer M, Briceno I, Camp NJ, Campbell A, Castelao JE, Chang-Claude J, Chanock SJ, Chenevix-Trench G, Collaborators N, Collée JM, Czene K, Dennis J, Dörk T, Eriksson M, Evans DG, Fasching PA, Figueroa J, Flyger H, Gabrielson M, Gago-Dominguez M, García-Closas M, Giles GG, Glendon G, Guénel P, Gündert M, Hadjisavvas A, Hahnen E, Hall P, Hamann U, Harkness EF, Hartman M, Hogervorst FBL, Hollestelle A, Hoppe R, Howell A, Investigators k, Investigators S, Jakubowska A, Jung A, Khusnutdinova E, Kim S-W, Ko Y-D, Kristensen VN, **Lakeman IMM**, Li J, Lindblom A, Loizidou MA, Lophatananon A, Lubiński J, Luccarini C, Madsen MJ, Mannermaa A, Manoochehri M, Margolin S, Mavroudis D, Milne RL, Taib NAM, Muir K, Nevanlinna H, Newman WG, Oosterwijk JC, Park SK, Peterlongo P, Radice P, Saloustros E, Sawyer EJ, Schmutzler RK, Shah M, Sim X, Southey MC, Surowy H, Suvanto M, Tomlinson I, Torres D, Truong T, Asperen CJv, Waltes R, Wang Q, Yang XR, Pharoah PDP, Schmidt MK, Benitez J, Vroling B, Dunning AM, Teo SH, Kvist A, Hoya Mdl, Devilee P, Spurdle AB, Vreeswijk MPG, Easton DF. Breast cancer risks associated with missense variants in breast cancer susceptibility genes. *Submited*

Curriculum Vitae

Inge Margaretha Maria Lakeman werd geboren op 6 januari 1987 in Hoorn en groeide op in Onderdijk. In 2005 behaalde zij haar VWO diploma aan het Martinus College in Grootebroek. In datzelfde jaar begon zij de studie Biomedische Wetenschappen aan de Universiteit van Amsterdam. Tijdens deze studie werd haar interesse voor de klinische genetica en wetenschappelijk onderzoek gewekt. In 2008 behaalde zij haar bachelor Biomedische Wetenschappen cum laude en begon zij de studie Geneeskunde aan de Vrije Universiteit (VU) Amsterdam. In het laatste jaar van haar studie heeft zij haar semi-arts stage gedaan op de afdeling Interne Geneeskunde van het Zaans Medisch Centrum en haar wetenschappelijke stage op de afdeling Klinische Genetica van het VU medisch centrum in de groep van dr. G. Pals. Tijdens deze stage deed zij onderzoek naar de expressie van PLS3 in vrouwen met osteoporose of milde osteogenesis imperfecta. Na afronding van haar master geneeskunde in 2014 is zij gaan werken als ANIOS bij de afdeling Klinische Genetica van het VU medisch centrum, met als aandachtsgebied oncogenetica. In 2015 startte zij met haar promotieonderzoek in de groep van prof. dr. P. Devilee bij de afdeling Humane Genetica van het Leids Universitair Medisch Centrum (LUMC), mede onder begeleiding van prof. dr. C.J. van Asperen van de afdeling Klinische Genetica van het LUMC, waarvan het resultaat staat beschreven in dit proefschrift. In het kader van de studies beschreven in dit proefschrift bezocht zij diverse congressen in het binnen- en buitenland waar ze meerdere mondelinge en posterpresentaties gaf, waaronder een voordracht op het European Human Genetics Conference in Gothenburg, Zweden. Tijdens haar promotieonderzoek heeft zij gedurende enkele maanden gewerkt in de groep van prof. dr. M.K. Schmidt in het Nederlands Kanker Instituut wat resulteerde in een publicatie beschreven in hoofdstuk 5 van dit proefschrift. Sinds 2020 combineert zij haar promotieonderzoek met het werk in de kliniek op de afdeling Klinische Genetica in het LUMC. In 2021 is zij gestart met de opleiding tot Klinisch Geneticus in het LUMC.

Dankwoord

Promoveren doe je gelukkig niet alleen. In de laatste woorden van dit proefschrift wil ik graag iedereen bedanken die hieraan heeft bijgedragen.

Allereerst alle betrokken patiënten en families voor het beschikbaar stellen van hun DNA en medische gegevens, wat dit onderzoek mogelijk heeft gemaakt.

Mijn promotoren Peter en Christi. Bedankt dat ik dit onderzoek onder jullie supervisie heb mogen uitvoeren. Jullie verschillende invalshoeken hebben bijgedragen aan dit mooie resultaat. Peter, je deur stond altijd voor mij open voor vragen of leuke discussies, je wist mijn Engelse teksten veel mooier te maken en je enthousiasme zorgde voor werkplezier. Christi, bedankt voor je betrokkenheid en klinische blik. Door jouw doelgerichtheid is dit proefschrift tot een mooi einde gekomen.

Mar, what should I do without your help on statistical topics. Thank you for your enthusiasm and for always being available to answer my questions.

Marjanka, bedankt dat ik onder jouw supervisie aan de CIMBA studie mocht werken en dat je ook daarna bij mijn onderzoek betrokken bent gebleven. Jouw kennis en positieve houding heb ik zeer gewaardeerd.

Van de vele personen die betrokken zijn geweest bij mijn onderzoek wil ik een paar in het bijzonder bedanken. Nandi, bedankt voor het voorbereiden van alle DNA samples, dat was een hele klus. Denise, bedankt voor alle HEBON informatie die je keer op keer met mij hebt gedeeld en voor je hulp bij de pilot studie. Dat we samen de enorme uitdaging van het verzamelen van de stambomen aan konden gaan maakte het een stuk leuker. Merel, bedankt voor je hulp met het verwerken van alle stambomen en klinische data maar ook zeker voor de gezelligheid als mijn kamergenoot.

Ik wil graag mijn collega's uit de Tumorgenetica groep bedanken, met name mijn kamergenoten Maaike en Ellen, voor de gezellige tijd die ik daar heb gehad, de leuke koffie/lunch momenten en de uitjes samen. Ook mijn nieuwe collega's bij de klinische genetica en mede AIOS die als geen ander weten wat het is om de opleiding tot klinisch geneticus met de afronding van je promotieonderzoek te combineren.

Naast mijn collega's hebben veel personen mij bijgestaan door te zorgen voor de nodige ontspanning naast het maken van dit proefschrift. Lieve vriendinnen, bedankt voor alle leuke afleiding in de vorm van etentjes, borrels, koffies en weekendjes weg. Louise, na 9

jaar samen studeren vond ik het meer dan terecht dat jij als paranimf hier naast mij mag staan. Wat ben ik blij dat wij nog steeds zulke goede vriendinnen zijn.

Wim, Mariët, Carmen en Arjen, wat fijn dat ik mij vanaf het begin af aan welkom heb gevoeld bij jullie met als bonus dat ik tante mag zijn van Sebe en Rena. Kim, de keuze voor jou als paranimf was niet moeilijk. Het voelt als thuiskomen bij jou en wat heb je me trots gemaakt als tante van Ilse, Jasper en Tygo. Robin, bedankt dat je altijd voor me klaar staat en mij de nodige dingen uit mijn (linker) handen neemt. David en Petra, ik vind het heel leuk dat jullie al lang onderdeel zijn van onze familie.

Lieve pap en mam, bedankt dat jullie er altijd voor mij zijn. Door jullie steun en vertrouwen was elke stap makkelijker te nemen. Lieve Mark, zal ik jou bedanken voor dit proefschrift of is het andersom? Hoe dan ook, ik ben het gelukkigst met jou en onze prachtige dochter Benthe naast mij.

APPENDIX

A

Argumenten voor centrale toetsing niet WMO-plichtig multicenter onderzoek

Inge M.M. Lakeman, Arts-onderzoeker, afdeling Humane Genetica, LUMC
Martine C. de Vries, Hoogleraar Medische Ethiek, LUMC
Evert-Ben van Veen, Jurist, MedLawconsult
Christi J. van Asperen, Hoogleraar afdelingshoofd Klinische Genetica, LUMC

De finale versie van dit artikel is geschreven in Februari 2018

Inleiding

In dit artikel willen we de gang van zaken bediscussiëren met betrekking tot toetsing door de Medisch Ethische Toetsingscommissies (METC's) van multicenter onderzoek binnen Nederland, dat niet onder de reikwijdte van de Wet medisch-wetenschappelijk onderzoek met mensen (*niet* WMO-plichtig) valt. Voor onderzoek dat wèl WMO-plichtig is, wordt bij multicenter onderzoek het onderzoeksprotocol op grond van de WMO door één centrale METC beoordeeld. Vervolgens beoordelen de Raden van Bestuur (RvB)/Directies van de deelnemende centra, de lokale uitvoerbaarheid.[1] Dit voorkomt de mogelijkheid van verschillen in de beoordeling van het onderzoeksprotocol. Bij *niet* WMO-plichtig multicenter onderzoek zouden de deelnemende centra één toetsingscommissie kunnen aanwijzen, maar in de praktijk komt hier niets van terecht. Het onderzoeksprotocol moet daarom in elke deelnemende instelling apart beoordeeld worden. Dit kan leiden tot verschillende beoordelingen en uitslagen van de METC's. Met name bij studies binnen families kan dit een nadeel zijn, indien familieleden in verschillende instellingen bekend zijn. Dit illustreren we aan de hand van een landelijke studie.

Voorbeeld van een landelijke studie: BRIDGES-NL

Een studie naar de klinische toepasbaarheid van nieuwe borstkankergenen is het BRIDGES (Breast cancer Risk after Diagnostic GEne Sequencing) project (bron: https://bridges-research.eu/). BRIDGES is een internationale Europese studie waarbij een gen-panel test, bestaande uit 35 genen geassocieerd met borstkanker, bij borstkankerpatiënten wordt uitgevoerd. Een aantal van deze genen is tevens geassocieerd met een verhoogd risico op andere tumoren. In totaal worden in het BRIDGES project 60.000 DNA samples geïncludeerd waarvan ongeveer 2800 DNA samples van Nederlandse familiaire borstkankerpatiënten (BRIDGES-NL). Deze Nederlandse patiënten zijn geïncludeerd vanuit de HEBON (Hereditair Borst- en eierstokkanker Onderzoek Nederland) studie (bron: www.hebon.nl).

Deelnemers van HEBON zijn geïncludeerd rond 2012, na counseling en genetische diagnostiek voor borst- en/of eierstokkanker in een van de negen klinisch genetische centra in Nederland (NKI/AvL, VUmc, AMC, UMCG, LUMC, UMCM, RUMC, EMC en UMCU). De deelnemers hebben een deelnemersverklaring ondertekend en een vragenlijst ingevuld over risicofactoren en familiegeschiedenis. In de brochure bij de deelnemersverklaring staat beschreven dat DNA opgevraagd kan worden voor wetenschappelijk onderzoek en dat in principe geen individuele terugkoppeling zal plaatsvinden, tenzij bevindingen worden gedaan van groot klinisch belang voor de gezondheid.[2] Dit HEBON informed consent werd door de onderzoekers van BRIDGES-NL in eerste instantie voldoende geacht voor deelname. Vanwege het feit dat terugkoppeling van het onderzoeksresultaat nodig kan zijn aan een klein aantal deelnemers (n=~10) en het onderzoeksresultaat niet alleen een verhoogd risico op borstkanker kan betekenen maar ook eventueel een verhoogd

risico op andere tumoren, was er twijfel over het volstaan met de deelnemersverklaring van HEBON. Besloten werd het BRIDGES-NL onderzoek door de lokale METC's te laten beoordelen, als een addendum van het eerder goedgekeurde HEBON onderzoek.

Beoordeling niet WMO-plichtig onderzoek

In het BRIDGES-NL onderzoek maken we gebruik van diagnostisch verkregen DNA samples en wordt de patiënt niet opnieuw benaderd voor aanvullend onderzoek, materiaal of vragen. Het onderzoek is daarom *niet* WMO-plichtig en hierbij geldt het lokale beleid van de RvB en daarmee van de METC in een deelnemend centrum.[3, 4] In een aantal centra wordt volstaan met een beoordeling of het onderzoek wel of niet WMO-plichtig is. Bij de meeste centra dient het onderzoeksprotocol ook aan de METC voorgelegd te worden voor een zorgvuldigheidstoets, waarbij gebruik wordt gemaakt van de Code Goed Gebruik, Code Goed Gedrag, de WGBO en van de Wet bescherming persoonsgegevens.[5] Volgens de Code Goed Gebruik is het verstrekken van informatie over het onderzoek noodzakelijk en toestemming van de persoon vereist indien lichaamsmateriaal gebruikt wordt dat (in)direct herleidbaar is tot de persoon.[3, 6] Indien bevindingen bij het onderzoek zeker te verwachten zijn, moet eventuele terugkoppeling hiervan en de manier van terugkoppeling besproken worden met de deelnemer.[6] Bij goedkeuring van een *niet* WMO-plichtig onderzoek zal een verklaring van geen bezwaar afgegeven worden door de lokale METC.[5]

Nadelen van lokale toetsing

Bij het BRIDGES-NL onderzoek in Nederland zijn negen verschillende centra en daarmee negen verschillende RvB's en METC's betrokken bij de goedkeuring van het addendum, met ieder zijn eigen toetsingsprocedures. De verschillen bestonden onder andere uit de vereiste documenten en de wijze van beoordeling. De verschillen in de toetsingsprocedures zorgen voor onnodig meer werk voor de onderzoekers en vertraging van het onderzoek. Eerder zijn de lokale verschillen in toetsingsprocedures beschreven voor WMO-plichtig onderzoek, waar dezelfde problemen aan het licht werden gebracht.[7] Naast deze nadelen kan ook de onderzoeker zelf het addendum niet indienen bij de METC in een deelnemend centrum. Het is nodig dat een lokaal persoon, in het geval van het BRIDGES-NL onderzoek de HEBON vertegenwoordiger van het deelnemend centrum, het addendum indient. De onderzoeker is daardoor afhankelijk van de beschikbaarheid van een lokaal persoon in een deelnemend centrum. Naast het feit dat de verschillen in toetsingsprocedures leiden tot vertraging van het onderzoek, is ook de kans op een niet unanieme beoordeling groter. Er vindt namelijk geen overleg plaats tussen de METC's van de verschillende centra.

Verschillende METC beoordelingen BRIDGES-NL

De beoordeling van het addendum over het BRIDGES-NL onderzoek leidde uiteindelijk tot vier verschillende uitslagen van de METC's (figuur 1 en tabel 1). In de praktijk betekent deze uitslag dat er verschil bestaat in de informatievoorziening aan de deelnemers van het BRIDGES-NL onderzoek en in de gevraagde toestemming. In dit onderzoek zijn meerdere familieleden uit één familie geïncludeerd, die bekend kunnen zijn in verschillende centra. Dit zal dan tevens leiden tot verschil in informatievoorziening en toestemming binnen één familie. Daarnaast kunnen de verschillende uitslagen van de METC's verwarring brengen bij de onderzoekers en collega's uit andere centra over hoe deelnemers nu goed benaderd en geïnformeerd moeten worden.

De toetsingsprocedures in alle negen centra samen waren tijdrovend vanwege de verschillen in de vereiste documenten en aantal personen die betrokken waren bij het indienen en beoordelen. Een gevolg was dat we niet alle DNA samples hebben kunnen includeren omdat de deadline op een gegeven moment verstreken was voor inclusie in het BRIDGES project.

Tabel 1. Uitslagen van de METC's over het BRIDGES-NL addendum van HEBON

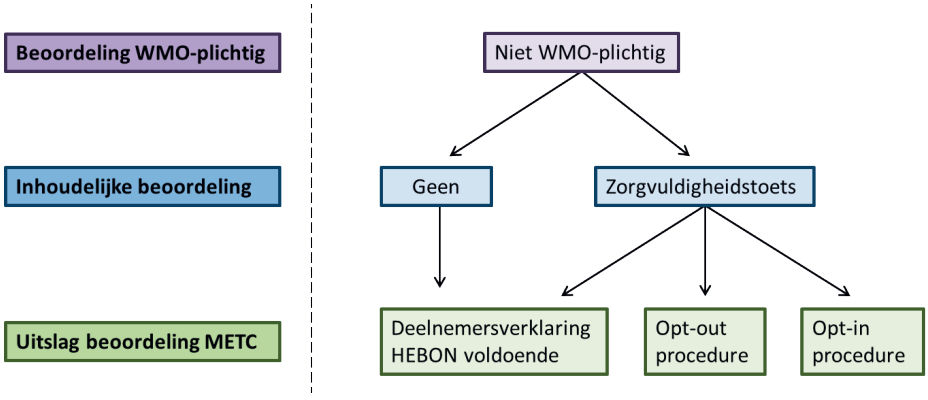
Procedure	Aantal centra	Aantal deelnemers benaderd	Weigeraar (opt-out) of geen reactie (opt-in)	Inclusie
Deelnemersverklaring voldoende	5	n.v.t.	n.v.t.	1363
Opt-out procedure	3	1156	25	1131
Opt-in procedure*	1	459	86	333**
Totaal	9	1615	111	2827

Opt-out procedure: Deelnemers informeren over het BRIDGES-NL onderzoek en daarbij de mogelijkheid bieden om zich terug te trekken uit het onderzoek.
Opt-in procedure: Deelnemers informeren over het BRIDGES-NL onderzoek waarbij (opnieuw) toestemming moet worden gegeven voor deelname aan BRIDGES-NL.
* Eenmalig een brief verstuurd naar de deelnemers
** I.v.m. de deadline voor het insturen van DNA samples zijn er 333 DNA samples geïncludeerd in plaats van 373 waarvoor toestemming is verkregen.

Eerder initiatief centrale toetsing

Voor *niet* WMO-plichtig onderzoek loopt inmiddels het project 'eenvormige toetsing', van COmmissie REgelgeving ONderzoek (COREON) naar aanleiding van vergelijkbare problemen rond de toetsing van *niet* WMO-plichtig onderzoek. [8, 9] Dit project is er op gericht om ook voor *niet* WMO-plichtig multicenter onderzoek een centrale toetsing mogelijk te maken, zoals voor WMO-plichtig onderzoek wordt gehanteerd. Door de industrie gesponsord en geïnitieerd *niet* WMO-plichtig onderzoek wordt inmiddels wél centraal getoetst. Dit is door het Ministerie van Volksgezondheid, Welzijn en Sport (VWS)

gesubsidieerd en vertegenwoordigt maar een zeer gering deel van het totaal aan *niet* WMO-plichtig onderzoek.[10] De subsidieaanvraag voor het project 'eenvormige toetsing', dat een veel groter deel van het *niet* WMO-plichtig onderzoek vertegenwoordigt, werd echter door het ministerie van VWS afgewezen. Het centraal toetsen van het overgrote deel van het *niet* WMO-plichtig onderzoek, kan niet van bovenaf worden opgelegd. Dit zal door onderling vertrouwen moeten plaatsvinden, bijvoorbeeld uitmondend in een convenant waarbij alle toetsingscommissies elkaars oordelen accepteren. Om dit te organiseren zijn noodzakelijke middelen vereist die op dit moment nog ontbreken. Het onderzoeksveld (COREON en BBMRI-NL(Biobanking and BioMolecular resources Research Infrastructure-The Netherlands)) heeft al veel in het project "eenvormige toetsing"geïnvesteerd. Zodra de additionele middelen beschikbaar komen, kan dit verder worden opgepakt.



Figuur 1. Verschillende beoordelingsprocedures en uitslagen METC
Opt-out procedure: Deelnemers informeren over het BRIDGES-NL onderzoek en daarbij de mogelijkheid bieden om zich terug te trekken uit het onderzoek.
Opt-in procedure: Deelnemers informeren over het BRIDGES-NL onderzoek waarbij (opnieuw) toestemming moet worden gegeven voor deelname aan BRIDGES-NL.

Conclusie

Het is duidelijk dat er noodzaak is voor centraal toetsen van onderzoek welke niet onder de reikwijdte van de WMO valt. Dit zal meervoudig toetsen voorkomen en de efficiëntie aanzienlijk verbeteren waardoor tegenstrijdige oordelen zoals bij het BRIDGES-NL onderzoek vermeden kunnen worden. Zeker door de toenemende mogelijkheden voor genetisch onderzoek binnen families verdeeld over heel Nederland, is het niet wenselijk dat deelnemers op verschillende wijzen benaderd worden voor hetzelfde onderzoek. Evenals voor WMO-plichtig onderzoek blijft de beoordeling van de privacy regelgeving en de toegang en opslag van data onder de verantwoordelijkheid van de lokale RvB in elk centrum. Een ander punt is dat er overeenstemming bereikt kan worden over de vereiste documenten voor toetsing bij de METC's. Door gebruik van universele documenten

zal de efficiëntie van de toestemmingsprocedure verder verbeteren. Het is een gemis dat centrale toetsing van *niet* WMO-plichtig multicenter onderzoek nog steeds niet in Nederland plaatsvindt. Veel tijd gaat verloren wat niet ten goede komt aan de kwaliteit van het onderzoek voor de patiënt.

Referenties

1. CCMO. Het verkrijgen van goedkeuring voor multicenteronderzoek. 2012. URL: www.ccmo.nl/attachments/files/brochure-nieuwe-procedure-multicenteronderzoek-5-11-2011.pdf.
2. HEBON. Deelnemersinformatie. URL: www.hebon.nl/bestanden/hebon-brochure.pdf.
3. CCMO, niet-WMO-onderoek. [Geraadpleegd in juni 2017]. URL: www.ccmo.nl/nl/niet-wmo-onderzoek.
4. CCMO, uw onderzoek: WMO-plichtig of niet [Geraadpleegd in juni 2017]. URL: www.ccmo.nl/nl/uw-onderzoek-wmo-plichtig-of-niet.
5. LUMC, niet-WMO onderzoek [Geraadpleegd in juni 2017].URL: www.lumc.nl/org/metc/toetsingsprocedures/niet-WMO-onderzoek/.
6. Federa. Verantwoord omgaan met lichaamsmateriaal ten behoeve van wetenschappelijk onderzoek. 2015. URL: www.federa.org/sites/default/files/images/codegoedgebruik_versiea4_juli_2015_beeldmerk_federa_en_coreon_corr_pag_4_jvds.pdf.
7. van der Stok EP, Huiskens J, Hemmes B, Grunhagen DJ, van Gulik TM, Verhoef C, et al. Lokale toestemmingsprocedures zetten een rem op RCT's. Nederlands tijdschrift voor geneeskunde. 2016;160(0):D148.
8. Eenvormige toetsing. nWMO-plichtig onderzoek [Geraadpleegd in november 2017]. URL: www.eenvormigetoetsing.nl/.
9. Federa. over COREON [Geraadpleegd in november 2017]. URL: www.federa.org/over-coreon.
10. nWMO studies. Toetsingskader niet WMO plichtig onderzoek [Geraadpleegd in november 2017]. URL: https://nwmostudies.nl/.