

## The associations between a polygenic score, reproductive and menstrual risk factors and breast cancer risk

Shaneda Warren Andersen · Amy Trentham-Dietz · Ronald E. Gangnon · John M. Hampton · Jonine D. Figueroa · Halcyon G. Skinner · Corinne D. Engelman · Barbara E. Klein · Linda J. Titus · Polly A. Newcomb

Received: 17 April 2013 / Accepted: 18 July 2013 / Published online: 27 July 2013  
© Springer Science+Business Media New York 2013

**Abstract** We evaluated whether 13 single nucleotide polymorphisms (SNPs) identified in genome-wide association studies interact with one another and with reproductive and menstrual risk factors in association with breast cancer risk. DNA samples and information on parity, breastfeeding, age at menarche, age at first birth, and age at menopause were collected through structured interviews from 1,484 breast cancer cases and 1,307 controls who participated in a population-based case–control study

conducted in three US states. A polygenic score was created as the sum of risk allele copies multiplied by the corresponding log odds estimate. Logistic regression was used to test the associations between SNPs, the score, reproductive and menstrual factors, and breast cancer risk. Nonlinearity of the score was assessed by the inclusion of a quadratic term for polygenic score. Interactions between the aforementioned variables were tested by including a cross-product term in models. We confirmed associations between rs13387042 (2q35), rs4973768 (*SLC4A7*), rs10941679 (5p12), rs2981582 (*FGFR2*), rs3817198 (*LSP1*), rs3803662 (*TOX3*), and rs6504950 (*STXBP4*) with breast cancer. Women in the score's highest quintile had 2.2-fold increased risk when compared to women in the lowest quintile (95 % confidence interval: 1.67–2.88). The quadratic polygenic score term was not significant in the model ( $p = 0.85$ ), suggesting that the established breast cancer loci are not associated with increased risk more than the sum of risk alleles. Modifications of menstrual and reproductive risk factors associations with breast cancer risk by polygenic score were not observed. Our results suggest that the interactions between breast cancer susceptibility loci and reproductive factors are not strong contributors to breast cancer risk.

S. Warren Andersen (✉) · A. Trentham-Dietz · J. M. Hampton · P. A. Newcomb  
University of Wisconsin Carbone Cancer Center, 610 Walnut St, WARF Room 307, Madison, WI 53726, USA  
e-mail: snandersen@uwalumni.com

S. Warren Andersen · A. Trentham-Dietz · R. E. Gangnon · H. G. Skinner · C. D. Engelman  
Department of Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI 53726, USA

J. D. Figueroa  
Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, MD 20892, USA

B. E. Klein  
Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI 53726, USA

L. J. Titus  
Norris Cotton Cancer Center, Dartmouth Medical School, Lebanon, NH 03756, USA

P. A. Newcomb  
Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

**Keywords** Epidemiology · Reproductive and menstrual factors · Breast cancer · Breast cancer susceptibility loci

### Abbreviations

CI Confidence interval  
GWAS Genome-wide association study  
MAF Minor allele frequency  
OR Odds ratio  
SNP Single nucleotide polymorphism

## Introduction

Genome-wide association studies (GWAS) and their follow-up studies have identified numerous single nucleotide polymorphisms (SNPs) with unknown biologic significance to be associated with breast cancer risk [1–7]. Researchers have begun to create composite genetic risk scores to investigate the polygenic manner of the genetic variants [8–11]. In a study by Reeves et al. [9], women in the highest quintile of a polygenic score that incorporated seven breast cancer susceptibility loci experienced a 40 % (95 % CI = 31–48 %) increased odds of breast cancer when compared to the middle quintile reference group. The association between the polygenic risk score and breast cancer risk was attenuated in participants diagnosed with estrogen receptor-negative tumors as opposed to estrogen receptor-positive tumors, providing evidence that some breast cancer susceptibility loci may be more strongly related to hormonally motivated tumors [9, 12].

Researchers hypothesize that certain genetic variants, in conjunction with reproductive and menstrual factors, are involved in hormonal pathways to influence breast cancer risk. Several studies have examined effect modification of reproductive and menstrual factors by GWAS-identified susceptibility loci and have found mainly null associations, [6, 13, 14] however, a number of modest interactions have been reported between reproductive and menstrual risk factors and breast cancer susceptibility loci, most notably with parity, age at menarche, and age at natural menopause [6, 13, 15–17]. For instance, an interaction between age at menarche and an established breast cancer susceptibility SNP, rs13387042 (2q35), was recently described in the literature. Women with older ages at menarche ( $\geq 13$  years) had an attenuated 12 % per-allele increased breast cancer risk compared to women with younger ages at menarche (22 % per-allele increase; *interaction p*-value = 0.04) [13]. Another study found the association between rs3817198 (*LSP1*) and breast cancer risk was stronger in women with more live births. The authors found a 4 % increased risk of breast cancer for every increase in the number of children and increase in the number of minor alleles [6]. A recent follow-up study which utilized data from more than 34,000 cases and 41,000 controls further confirmed this finding [17]. A separate study found established reproductive and menstrual factors and a polygenic score contributed independently to breast cancer risk [11]. The authors noted the association between their polygenic score and breast cancer risk did not differ by menopausal status of the participants. However, the researchers did not assess whether other reproductive and menstrual factors modified the association between the polygenic score and breast cancer risk [11]. Moreover, most previous studies have not systematically examined whether a composite risk

score modifies the associations between established reproductive and menstrual factors with breast cancer risk. We investigated the associations between established breast cancer susceptibility loci with one another and reproductive and menstrual factors in association with breast cancer risk in a population-based case–control study.

## Materials and methods

### Study sample

Data were collected from the Three State Study, a previously described population-based breast cancer case–control study [18–20]. Participants were selected from English-speaking females residing in Massachusetts (excluding metropolitan Boston), New Hampshire, and Wisconsin. Cases included in this analysis were women age 20–69 with an incident invasive breast cancer reported to each state’s cancer registry between 1995 and 2000. Community controls were randomly selected in each state from lists of licensed drivers (<age 65) and lists of Medicare beneficiaries ( $\geq$ age 65). Controls were frequency-matched to approximate the age distribution of the cases within 5-year age strata. Participants gave informed consent during study enrollment. This study was conducted under the approval of the University of Wisconsin Health Sciences Institutional Review Board.

### Risk factor information

Telephone interviews were used to obtain information on known and suspected risk factors for breast cancer including demographics, first degree family history of breast cancer, and hormonal exposures. Participant interviews were conducted on average 1 year after a specified reference date, which was defined as the date of cancer diagnosis for the cases. A comparable reference date for control participants was calculated based on their 5-year age strata and date of interview [20]. Among eligible participants, approximately 80 % of cases and 76 % of controls completed the interview.

### DNA extraction and genotyping

Selected participants were asked to donate a buccal cell sample for genetic analyses using an oral rinse protocol. Participants interviewed between the years 2000 and 2001, who provided a buccal cell sample, are included in the present analysis. 70 % of these interviewed cases and 61 % of controls agreed to donate a buccal sample. Participants who chose not to provide a buccal cell sample were similar in age, percentage with a family history of breast cancer, and

other established risk factors for breast cancer. To reduce the possibility of population stratification and maintain a study sample with ancestry similar to the GWAS and their follow-up studies, all analyses were limited to participants self-identified as White/Caucasian in race (95.1 % of participants). Samples were sent to a National Cancer Institute-affiliated laboratory for DNA extraction and storage conducted according to previously described protocols [18]. DNA was quantitated from frozen aliquots and plated for the genotyping assays. Significant results from previous studies were used to identify 13 candidate SNPs for the analysis: rs4973768 (*SLC4A7*), rs10941679 (5p12), rs2981582 (*FGFR2*), rs3817198 (*LSP1*), rs3803662 (16q12/*LOC643714/TOX3*), rs13281615 (8q24), rs11249433 (1p11), rs889312 (*MAP3K1*), rs2046210 (6q25), rs17468277 (*ALS2CR12/CASP8*), rs10483813 (*RAD51B*), rs13387042 (2q35), and rs6504950 (*STXBP4*) [1–6, 21, 22]. Genotyping was conducted using the Taqman nuclease assay (Taqman<sup>®</sup>) with reagents designed by Applied Biosystems (<http://www.appliedbiosystems.com/>) as Assays-by-Design<sup>™</sup> and genotyping performed using the ABI PRISM 7900HT, 7700 or 7500 Sequence Detection Systems according to the manufacturer's instructions. Quality control measures were taken to remove poor quality genotype data. SNPs missing >20 % of values or individual participants with a call rate <80 % for genotypic data were excluded from the analysis. All 13 SNPs passed quality control measures. 358 participants were removed from genetic analyses due to missing data for a total of 1,484 breast cancer cases and 1,307 community controls.

#### Covariate definitions

All reproductive and menstrual variables were first coded continuously and secondarily categorized into subgroups based on hypothetical biologic differences in risk. Menarche was categorized into tertiles to represent early, average, and late age at menarche (<12, 12–13, ≥14). Age at first birth was coded only among parous women and was coded according to frequently-used cutoffs (<20, 20–24, 25–29, ≥30). Parity was categorized as nulliparous and then in tertiles, among women that had ever been pregnant (1–2, 3, ≥4 live births). Participants were considered postmenopausal if they reported their menstrual cycles had stopped for at least the last 6 months prior to the reference date. The menopausal participants were categorized into two groups: participants with natural menopause or a second group defined as menopause due to other causes which included women whose menstrual periods stopped because they underwent bilateral oophorectomy, stopped using hormonal contraceptives, or had an unknown cause of menopause. Family history of breast cancer was defined as

having at least one first degree relative (e.g., mother, sister, and daughter) with a breast cancer diagnosis.

#### Statistical analyses

Hardy–Weinberg equilibrium was tested among controls by using Chi squared tests to compare the observed to expected genotype frequencies. Odds ratios (OR) and 95 % confidence intervals (CI) were calculated using logistic regression to assess the association between each SNP and breast cancer risk under an additive genetic model with respect to the minor allele. All statistical models included a term for age and state of residence. Associations between established risk factors and breast cancer risk were also calculated. In order to evaluate the comparability of our point estimates to previous studies, we compared Three State Study breast cancer susceptibility loci point estimates to the estimates reported in the published GWAS or GWAS follow-up study by normal standardization. Statistical analysis was conducted in SAS software (Cary, NC 9.1).

#### Polygenic risk score

A composite risk score was created to assess the polygenic contribution of breast cancer susceptibility loci. All SNPs were coded as a count of the number of risk alleles. A stepwise selection procedure with a stay and entry criteria of 0.1 was used to identify SNPs most strongly associated with breast cancer risk (SAS software version 9.1). A weighted risk score was then formed as the sum of the number of risk allele copies of the selected SNPs multiplied by the corresponding log odds estimate. Nonlinearity of the score was assessed by the inclusion of a quadratic term, and interactions between the score and established reproductive and menstrual factors were tested by including a cross-product term in statistical models. In order to capture the polygenic risk score's association with breast cancer risk when established reproductive and menstrual exposures were also considered, an additional model was analyzed by including the following variables selected a priori: age, state of residence, age at menarche, age at first full-term birth, parity, ever breastfeeding, and age at natural menopause. A third model was analyzed which included the aforementioned variables as well as a term for the presence of family history of breast cancer.

#### Reproductive and menstrual factor effect modification

Multivariate models were calculated to evaluate effect modification of the associations between reproductive and menstrual exposures with breast cancer risk by including cross-product terms combining the exposure of interest

multiplied by the polygenic score. The reproductive and menstrual risk factors assessed in this study are as follows: age at menarche, age at first full-term birth, parity, ever breastfeeding, and age at natural menopause. To further elucidate how risks differ by combinations of genotypes and reproductive and menstrual patterns, stratified odds ratios were calculated for the associations between reproductive or menstrual factors and breast cancer risk stratified by the polygenic score.

## Results

For no SNP, there was evidence for departure from Hardy–Weinberg Equilibrium ( $p$ -values  $> 0.05$ ). We found no statistically significant differences in the magnitude of the association between the calculated Three State Study odds ratios and the odds ratios reported by the GWAS or follow-up studies for the association between the 13 loci and breast cancer risk (Table 1). We confirmed previously reported associations between seven breast cancer susceptibility loci and invasive breast cancer risk: rs13387042 (2q35), rs6504950 (*STXBP4*), rs4973768 (*SLC4A7*), rs10941679 (5p12), rs2981582 (*FGFR2*), rs3817198 (*LSP1*), and rs3803662 (*TOX3*). The range of estimated increase in breast cancer risk per increase in risk alleles was 11–22 % with SNP rs2981582 (*FGFR2*) showing the strongest association with breast cancer risk in this study; the minor allele was associated with a 22 % increase in breast cancer risk (95 % CI 8–38 %).

### Polygenic risk score results

A selection procedure identified seven of the 13 SNPs (rs13387042, rs4973768, rs10941679, rs2981582, rs3817198, rs3803662, and rs6504950) to include in the polygenic risk score. The range of the number of risk alleles present in the risk score was similar in cases and controls, although the distribution in cases skewed toward higher numbers of risk alleles; the range of risk alleles was 1–12 (mean = 5.86) in controls and 2–12 (mean = 6.35) in cases. Women in the highest quintile of the score had a 2.2-fold increased breast cancer risk when compared to women in the lowest quintile (95 % CI 1.67–2.88). Women in the third and fourth quintiles also had an increased risk (OR = 1.52, 95 % CI 1.15–2.02; OR = 1.50, 95 % CI 1.13–1.98, respectively) (Table 2). A quadratic polygenic risk score term was added to the statistical model to assess nonlinearity and was not statistically significant ( $p$ -value = 0.85). Polygenic risk score models adjusted for reproductive and menstrual exposures did not materially change the composite point estimate. Moreover, results were similar when an additional term for family history was added to the model (Table 2).

Effect modification results for individual SNPs, the polygenic score, and reproductive and menstrual factors

We conducted 21 pairwise interaction tests among the seven significant SNPs (rs13387042, rs4973768, rs10941679, rs2981582, rs3817198, rs3803662, and rs6504950), and did not observe strong evidence of interactions in their associations with breast cancer risk (19  $p$ -values  $> 0.05$ ). Potential effect modification of rs13387042 by rs4973798 (interaction  $p$ -value = 0.02) and rs10941679 by rs3803662 (interaction  $p$ -value = 0.03) was noted. We also evaluated whether breast cancer susceptibility loci modified the associations between reproductive and menstrual risk factors with breast cancer risk. Sample distributions of these hormonal exposures are located in Table 3. Cases were more likely to have a first degree family history of breast cancer, earlier age at menarche, fewer children, and later age at menopause than controls. Effect modification of the associations between reproductive or menstrual factors and breast cancer risk by the polygenic score were not observed (all interaction  $p$  values  $> 0.05$ ) with the exception of age at natural menopause where there was a weak interaction detected ( $p$  value = 0.09, result not shown). The deleterious association between later age at natural menopause and breast cancer risk was more apparent in women with lower polygenic score values.

## Discussion

We genotyped 13 breast cancer susceptibility loci identified from previous genetic epidemiology studies to examine how these loci interact with each other, and reproductive and menstrual risk factors in association with breast cancer risk. Of the candidate SNPs, seven were confirmed for an association with breast cancer risk including rs2981582 in the fibroblast growth factor receptor 2 gene, the SNP most strongly associated with breast cancer risk in this population. SNP rs2981582 and the other six significant SNPs have also been confirmed as breast cancer loci in a number of study populations and ethnic subgroups [1, 6, 13, 23].

We examined the possibility that when multiple risk alleles are found in conjunction with one another their association with breast cancer risk may be non-additive. Our polygenic score indicated a linear association with increased breast cancer risk, and risk was more than doubled for women in the highest risk quintile compared to women in the lowest. We found no statistically significant differences between our odds ratio estimates and those previously reported in the literature. The comparability of the estimates supported using the point estimates from the current study to create our polygenic risk score. Previous

**Table 1** Per-allele odds ratios and 95 % confidence intervals for the association between breast cancer susceptibility loci and breast cancer risk

Chromosomal location/Gene	SNP	Risk Allele	Original OR (95% CI) <sup>a</sup>	Citation	RAF	ThreeState Study OR (95% CI) <sup>c</sup>	P <sub>TREND</sub>
10q26/ <i>FGFR2</i>	rs2981582	T	1.26 (1.23-1.30)	Easton[1]	0.40	1.22 (1.08-1.38)	0.001
16q12/ <i>TNRC9/TOX3</i>	rs3803662	T	1.20 (1.16-1.24)	Easton[1]	0.28	1.22 (1.07-1.38)	0.003
5p12	rs10941679	G	1.27 (1.19-1.35)	Stacey[21]	0.27	1.20 (1.05-1.37)	0.01
17q23/ <i>STXBPA</i>	rs6504950	G	0.92 (0.86-0.99) <sup>b</sup>	Ahmed[4]	0.73	1.20 (1.06-1.37)	0.01
3p24/ <i>SLC4A7</i>	rs4973768	T	1.11 (1.08-1.13)	Ahmed[4]	0.50	1.19 (1.06-1.34)	0.004
11p15/ <i>LSP1</i>	rs3817198	C	1.12 (1.05-1.19)	Easton[1]	0.32	1.16 (1.02-1.31)	0.02
2q35	rs13387042	T	1.20 (1.14-1.26)	Stacey[2]	0.54	1.12 (1.00-1.26)	0.04
8q24	rs13281615	G	1.08 (1.05-1.11)	Easton[1]	0.42	1.07 (0.95-1.20)	0.23
1p11	rs11249433	C	1.08 (1.02-1.15)	Thomas[3]	0.43	1.06 (0.95-1.19)	0.31
5q11/ <i>MAP3K1</i>	rs889312	C	1.13 (1.10-1.16)	Easton[1]	0.29	1.06 (0.93-1.20)	0.43
14q24/ <i>RAD51B</i>	rs10483813	T	0.92 (0.89-0.94) <sup>b</sup>	Figuroa[22]	0.79	1.02 (0.89-1.17)	0.73
6q25	rs2046210	T	1.15 (1.03-1.28)	Zheng[5]	0.37	1.00 (0.90-1.13)	0.80
2q33/ <i>CASP8</i>	rs17468277	A	0.94 (0.91-0.98)	Milne[6]	0.12	1.00 (0.84-1.19)	0.99

Table 1 displays odds ratios and 95 % confidence intervals for the associations between 13 breast cancer susceptibility loci identified through genome-wide association and follow-up studies. The x-axis of the figure displays the per-risk-allele odds ratio for each SNP. The horizontal lines represent 95 % confidence limits in the Three State Study

OR indicates odds ratio of the risk allele, RAF risk allele frequency, CI confidence interval

<sup>a</sup> For comparison of study results this column displays the published point estimate and 95 % confidence limit from the original genome-wide association or follow-up study

<sup>b</sup> The original reported rs6504950 odds ratio estimated the per-allele effect for the A allele. In the current study the per-A-allele odds ratio was 0.83 (95 % CI = 0.73–95). The original reported rs10483813 odds ratio estimated the per-allele effect for the A allele. In the current study the per-A-allele odds ratio was 0.98 (95 % CI = 0.85–1.12)

<sup>c</sup> Odds ratios are adjusted for age and state of residence

**Table 2** Odds ratios and 95 % confidence intervals for the association between a polygenic risk score a and breast cancer risk

Risk score quintile	Cases	Controls	OR <sup>b</sup> (95 % CI)	OR <sup>c</sup> (95 % CI)	OR <sup>d</sup> (95 % CI)	P <sub>Trend</sub>
1st	161	204	1 (ref.)	1 (ref.)	1 (ref.)	
2nd	178	204	1.14 (0.85–1.52)	1.44 (0.89–2.34)	1.12 (0.83–1.50)	
3rd	246	204	1.52 (1.15–2.02)	1.81 (1.13–2.90)	1.52 (1.14–2.02)	
4th	241	204	1.50 (1.13–1.98)	1.59 (0.99–2.53)	1.46 (1.10–1.95)	
5th	348	204	2.20 (1.67–2.88)	2.22 (1.42–3.49)	2.16 (1.64–2.85)	<0.0001

OR odds ratio, CI confidence interval, ref. reference

<sup>a</sup> Composite SNP risk score includes rs13387042, rs4973768, rs10941679, rs2981582, rs3817198, rs3803662 and rs6504950

<sup>b</sup> Odds ratios adjusted for age and state

<sup>c</sup> Odds ratios adjusted for age, state, age at menarche, age at first birth, parity, ever breastfeeding and age at natural menopause

<sup>d</sup> Odds ratios adjusted for age, state, age at menarche, age at first birth, parity, ever breastfeeding, age at natural menopause and family history of breast cancer

studies have found corresponding magnitudes of association between polygenic risk scores and breast cancer risk. Harlid et al. created a polygenic score using ten GWAS-identified SNPs, four of which are included in the risk score from the current study [rs2981582 (10q26), rs3803662

(16q12/*TOX3*), rs3817198 (11p15/*LSP1*), and rs13387042 (2q35)]. In their study, the OR was 2.12 (95 % CI 1.80–2.50) for women with the maximum number of risk alleles compared to those with the lowest number of risk alleles [8]. Reeves et al. genotyped the same ten SNPs as

**Table 3** Reproductive and menstrual characteristics by case status

Characteristic	Cases <i>N</i> (%) ( <i>N</i> = 1484)	Controls <i>N</i> (%) ( <i>N</i> = 1307)	OR (95 % CI) <sup>a</sup>
Age (median, standard deviation)	54.5 (9.0)	54.5 (9.4)	
Family history of breast cancer <sup>b</sup>			
No	1145 (77.2)	1108 (84.8)	1 (ref.)
Yes	303 (20.4)	175 (13.4)	1.48 (1.25–1.76)
Age at menarche (years)			
<12	317 (21.4)	255 (19.5)	1 (ref.)
12–13	799 (53.9)	668 (51.1)	0.96 (0.79–1.17)
≥14	350 (23.6)	374 (28.7)	0.75 (0.60–0.94)
Age at first full-term birth (years) <sup>c</sup>			
<20	213 (14.4)	211 (16.1)	1 (ref.)
20–24	609 (41.0)	536 (41.0)	1.12 (0.89–1.40)
25–29	316 (21.3)	266 (20.4)	1.23 (0.95–1.58)
≥30	143 (9.6)	118 (9.0)	1.26 (0.92–1.73)
Parity			
Nulliparous	199 (13.4)	171 (13.1)	1.01 (0.80–1.28)
1–2	656 (44.3)	565 (43.2)	1 (ref.)
3	324 (21.8)	270 (20.7)	1.00 (0.82–1.22)
≥4	303 (20.4)	298 (22.8)	0.82 (0.66–1.01)
Lactation (months) <sup>c</sup>			
Never	679 (52.8)	585 (47.3)	1 (ref.)
Ever	602 (46.8)	544 (44.0)	0.93 (0.79–1.10)
≤3	171 (11.5)	158 (12.1)	0.90 (0.70–1.15)
4–12	207 (14.0)	202 (15.5)	0.89 (0.70–1.12)
13–23	125 (8.4)	104 (8.0)	1.02 (0.76–1.36)
≥24	99 (6.8)	81 (6.2)	1.03 (0.75–1.43)
Menopause			
Premenopausal	565 (38.1)	470 (36.0)	1 (ref.)
Postmenopausal	822 (55.4)	742 (56.8)	0.75 (0.58–0.98)
Age at menopause (years)			
<45	123 (14.9)	130 (17.5)	1 (ref.)
45–49	171 (20.8)	173 (23.3)	1.04 (0.75–1.44)
50–54	255 (31.0)	220 (29.6)	1.22 (0.89–1.66)
≥55	91 (11.1)	59 (8.0)	1.60 (1.05–2.43)

OR indicates odds ratio, CI confidence interval, ref reference

<sup>a</sup> Odds ratios adjusted for age and state of residence

<sup>b</sup> Family history is defined as a first degree family member having been diagnosed with breast cancer

<sup>c</sup> Calculated only among parous women

the Harlid group and similarly found a twofold increase in risk when comparing the top and bottom quintiles of their polygenic score [9]. In the present study, we conclude that the increased breast cancer risk in women with larger risk scores was attributed to independent associations of the SNPs and not due to a synergistic increase in risk. Analogous to our results, Reeves et al. also found that the breast

cancer susceptibility loci were independently associated with breast cancer risk [9].

Only recently have researchers explored the possibility that breast cancer susceptibility loci in combination with reproductive and menstrual factors may increase breast cancer risk. We found that the polygenic score's association with breast cancer risk was separate from established

reproductive and menstrual factors as the association between the polygenic risk score and breast cancer risk did not materially change when reproductive and menstrual factors simultaneously considered. A study of breast cancer risk within the Women's Health Initiative Clinical Trial by Mealiffe et al. [10] assessed whether predictions of breast cancer risk could be improved by adding a polygenic score to the Gail risk model, which incorporates clinical and personal risk factor information, including age at menarche and age at first birth. Similar to our findings, Mealiffe et al. [10] found their polygenic score and the Gail model contributed separately to breast cancer risk. When their polygenic risk was added to statistical models which included the Gail model, the area under the curve increased from 0.557 to 0.594 ( $p$ -value <0.001).

This study has several strengths including comprehensive information on reproductive and menstrual factors obtained on a population-based sample with high participation rates. Previous studies have shown that women report reproductive and menstrual events with high accuracy [24] suggesting that our hormonal risk factor data should be reliability recorded. Additionally, previous investigators have not systematically evaluated whether a polygenic risk score modifies the associations between reproductive and menstrual factors with breast cancer risk, as we have done in this analysis. There are also certain considerations to be noted for this study's interpretation. Only a subset of established breast cancer susceptibility loci were evaluated in this study, consequently, loci important to the polygenic portion of breast cancer risk have not been included in the risk score leaving part of the genetic component of breast cancer risk unidentified. It is possible that a polygenic score which includes a more comprehensive set of loci may have a stronger association with invasive breast cancer risk than the risk score calculated here. We did not have information on tumor receptor status, and were unable to stratify breast cancer cases by many of the tumor characteristics known to be influenced by hormones. In summary, women with a higher risk score for seven established breast cancer loci were at an increased breast cancer risk compared to women with a lower polygenic score. Evidence from this study suggests that these loci are independently associated with breast cancer risk. Our polygenic score did not materially affect the associations between reproductive and menstrual risk factors with breast cancer risk.

**Acknowledgments** The authors are grateful to Dr. Kathleen M Egan for her input and advice over the course of the study. The authors have no conflicts of interest to disclose. This work was supported by the National Institutes of Health Intramural Research funds and grants (R01CA47147, R01CA47305, R01CA69664, U10EY006594), and by the Department of Defense Breast Cancer Research Program (W81XWH-11-1-0047).

## References

- Easton DF, Pooley KA, Dunning AM et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447:1087–1093. doi:10.1038/nature05887
- Stacey SN, Manolescu A, Sulem P et al (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 39:865–869. doi:10.1038/ng2064
- Thomas G, Jacobs KB, Kraft P et al (2009) A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* 41:579–584. doi:10.1038/ng.353
- Ahmed S, Thomas G, Ghoussaini M et al (2009) Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 41:585–590. doi:10.1038/ng.354
- Zheng W, Long J, Gao Y-T et al (2009) Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 41:324–328. doi:10.1038/ng.318
- Milne RL, Gaudet MM, Spurdle AB et al (2010) Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the breast cancer association consortium: a combined case-control study. *Breast Cancer Res Bcr* 12:R110. doi:10.1186/bcr2797
- Michailidou K, Hall P, Gonzalez-Neira A et al (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45:353–361. doi:10.1038/ng.2563
- Harlid S, Ivarsson MIL, Butt S et al (2012) Combined effect of low-penetrant SNPs on breast cancer risk. *Br J Cancer* 106:389–396. doi:10.1038/bjc.2011.461
- Reeves GK, Travis RC, Green J et al (2010) Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. *JAMA, J Am Med Assoc* 304:426–434. doi:10.1001/jama.2010.1042
- Mealiffe ME, Stokowski RP, Rhees BK et al (2010) Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst* 102:1618–1627. doi:10.1093/jnci/djq388
- Sueta A, Ito H, Kawase T et al (2012) A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. *Breast Cancer Res Treat* 132:711–721. doi:10.1007/s10549-011-1904-5
- Broeks A, Schmidt MK, Sherman ME et al (2011) Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum Mol Genet* 20:3289–3303. doi:10.1093/hmg/ddr228
- Travis RC, Reeves GK, Green J et al (2010) Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. *Lancet* 375:2143–2151. doi:10.1016/S0140-6736(10)60636-8
- Butt S, Harlid S, Borgquist S et al (2012) Genetic predisposition, parity, age at first childbirth and risk for breast cancer. *Bmc Res Notes* 5:414. doi:10.1186/1756-0500-5-414
- Campa D, Kaaks R, Le Marchand L et al (2011) Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst* 103:1252–1263. doi:10.1093/jnci/djr265
- Rebbeck TR, DeMichele A, Tran TV et al (2009) Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. *Carcinogenesis* 30:269–274. doi:10.1093/carcin/bgn247

17. Nickels S, Truong T, Hein R et al (2013) Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. *PLoS Genet* 9:e1003284. doi:[10.1371/journal.pgen.1003284](https://doi.org/10.1371/journal.pgen.1003284)
18. García-Closas M, Egan KM, Abruzzo J et al (2001) Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 10:687–696
19. García-Closas M, Egan KM, Newcomb PA et al (2006) Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. *Hum Genet* 119:376–388. doi:[10.1007/s00439-006-0135-z](https://doi.org/10.1007/s00439-006-0135-z)
20. Sprague BL, Trentham-Dietz A, Garcia-Closas M et al (2007) Genetic variation in TP53 and risk of breast cancer in a population-based case control study. *Carcinogenesis* 28:1680–1686. doi:[10.1093/carcin/bgm097](https://doi.org/10.1093/carcin/bgm097)
21. Stacey SN, Manolescu A, Sulem P et al (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 40:703–706. doi:[10.1038/ng.131](https://doi.org/10.1038/ng.131)
22. Figueroa JD, Garcia-Closas M, Humphreys M et al (2011) Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. *Hum Mol Genet* 20:4693–4706. doi:[10.1093/hmg/ddr368](https://doi.org/10.1093/hmg/ddr368)
23. Long J, Shu X-O, Cai Q et al (2010) Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 19:2357–2365. doi:[10.1158/1055-9965.EPI-10-0054](https://doi.org/10.1158/1055-9965.EPI-10-0054)
24. Bean JA, Leeper JD, Wallace RB et al (1979) Variations in the reporting of menstrual histories. *Am J Epidemiol* 109:181–185