Article:

Rudolf, A., Song, M., Brook, M.N. et al. (48 more authors) (2018) Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. International Journal of Epidemiology, 47 (2). pp. 526-536. ISSN 0300-5771

https://doi.org/10.1093/ije/dyx242

OXFORD
UNIVERSITY PRESS
International Journal of Epidemiology

Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium

Journal:	*International Journal of Epidemiology*
Manuscript ID	IJE-2017-04-0409.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	22-Sep-2017
Complete List of Authors:	Rudolph, Anja; Deutsches Krebsforschungszentrum Song, Minsun; Sookmyung Women's University Brook, Mark; Institute of Cancer Research Milne, Roger; The University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health; Cancer Council Victoria, Cancer Epidemiology Centre Mavaddat, Nasim; University of Cambridge Michailidou, Kyriaki; University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care; The Cyprus Institute of Neurology and Genetics, Department of Electron Microscopy/Molecular Pathology Bolla, Manjeet; University of Cambridge, Centre for Cancer Genetic Epidemiology Wang, Qin; University of Cambridge, Centre for Cancer Genetic Epidemiology Dennis, Joe; University of Cambridge, Centre for Cancer Genetic Epidemiology Wilcox, Amber; National Cancer Institute Division of Cancer Epidemiology and Genetics Hopper, John; <none>, Southey, Melissa; The University of Melbourne, Keeman, Renske; Nederlands Kanker Instituut - Antoni van Leeuwenhoek Ziekenhuis Fasching, Peter; University of California at Los Angeles, David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, ; University Hospital Erlangen, Department of Gynecology and Obstetrics, Friedrich-Alexander-University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen Nuremberg, Beckmann, Matthias; University Hospital Erlangen, Department of Gynecology and Obstetrics, Friedrich-Alexander-University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen Nuremberg, Gago-Dominguez, Manuela; Complejo Hospitalario Universitario de Santiago de Compostela Castelao, Jose; Complexo Hospitalario Universitario de Vigo Guenel, Pascal; INSERM, Truong, Thérèse; INSERM

Bojesen, Stig; Herlev Hospital, Department of Clinical Biochemistry
Flyger, Henrik; Rigshospitalet
Brenner, Hermann; Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany,
Arndt, Volker; Deutsches Krebsforschungszentrum
Brauch, Hiltrud; Doktor Margarete Fischer-Bosch-Institut fur Klinische Pharmakologie
Brüning, Thomas; IPA,
Mannermaa, Arto; Ita-Suomen yliopisto Kuopion kampus
Kosma, Veli-Matti; Ita-Suomen yliopisto Kuopion kampus
Lambrechts, Diether; VIB Vesalius Research Center
Keupers, Machteld; Universitaire Ziekenhuizen Leuven
Couch, Fergus; Mayo Clinic Minnesota
Vachon, Celine; Mayo Clinic Minnesota
Giles, Graham; Cancer Council Victoria, Cancer Epidemiology Centre
MacInnis, Robert; University of Melbourne, Centre for Molecular, Environmental, Genetic & Analytic (MEGA) Epidemiology in the School of Population and Global Health; Cancer Council Victoria, Cancer Epidemiology Centre
Figueroa, Jonine; University of Edinburgh Medical School
Brinton, Louise; National Cancer Institute, Division of Cancer Epidemiology & Genetics, Hormonal and Reproductive Epidemiology Branch
Czene, Kamila; Karolinska Institutet,
Brand, Judith; University Medical Center Utrecht, Julius Center for Health Sciences and Primary Care
Gabrielson, Marike; Karolinska Institutet, Medical epidemiology and biostatistics
Humphreys, Keith; Karolinska Institutet
Cox, Angela; University of Sheffield
Cross, SS; <none>,
Dunning, Alison; University of Cambridge
Orr, Nicholas; Institute of Cancer Research
Swerdlow, Anthony; Institute of Cancer Research, Section of Epidemiology
Hall, Per; Karolinska Institutet, Department of Medical Epidemiology and Biostatistics
Pharoah, Paul; Centre for Cancer Genetic Epidemiology, University of Cambridge,
Schmidt, Marjanka; Nederlands Kanker Instituut - Antoni van Leeuwenhoek Ziekenhuis
Easton, Douglas; University of Cambridge, Centre for Cancer Genetic Epidemiology
Chatterjee, Nilanjan; Johns Hopkins University Bloomberg School of Public Health
Chang-Claude, Jenny; German Cancer Research Center (DKFZ), Division of Cancer Epidemiology
Garcia-Closas, Montserrat; National Cancer Institute,

Key Words: breast cancer, genetic susceptibility, gene-environment interactions, risk prediction, epidemiology

Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium

Anja Rudolph,[1,2] Minsun Song,[3] Mark N. Brook,[4] Roger L. Milne,[5,6] Nasim Mavaddat,[7] Kyriaki Michailidou,[7,8] Manjeet K. Bolla,[7] Qin Wang,[7] Joe Dennis,[7] Amber N. Wilcox,[9] John L. Hopper,[6] Melissa C. Southey,[10] Renske Keeman,[11] Peter A. Fasching,[12,13] Matthias W. Beckmann,[12] Manuela Gago-Dominguez,[14,15] Jose E. Castelao,[16] Pascal Guénel,[17] Thérèse Truong,[17] Stig E. Bojesen,[18-20] Henrik Flyger,[21] Hermann Brenner,[22-24] Volker Arndt,[22] Hiltrud Brauch,[24-26] Thomas Brüning,[27] Arto Mannermaa,[28-30] Veli-Matti Kosma,[28-30] Diether Lambrechts,[31,32] Machteld Keupers,[33] Fergus J. Couch,[34] Celine Vachon,[35] Graham G. Giles,[5,6] Robert J. MacInnis,[5,6] Jonine Figueroa,[9,36] Louise Brinton,[9] Kamila Czene,[37] Judith S. Brand,[37] Marike Gabrielson,[37] Keith Humphreys,[37] Angela Cox,[38] Simon S. Cross,[39] Alison M. Dunning,[40] Nick Orr,[41] Anthony Swerdlow,[4,41] Per Hall,[37] Paul D.P. Pharoah,[7,40] Marjanka K. Schmidt,[11,42] Douglas F. Easton,[7,40] Nilanjan Chatterjee,[9,43,44] Jenny Chang-Claude,[1,45]† Montserrat García-Closas[9]†*

[1]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, [2]Real World Insights, CESE, QuintilesIMS, Frankfurt, Germany, [3]Department of Statistics, Sookmyung Women's University, Korea, [4]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK, [5]Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia, [6]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global health, The University of Melbourne, Melbourne, Australia, [7]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, [8]Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, [9]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA, [10]Department of Pathology, The University of Melbourne, Melbourne, Australia, [11]Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, [12]Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany, [13]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, USA, [14]Genomic Medicine Group, Galician Foundation of Genomic Medicine, Complejo Hospitalario Universitario de Santiago, SERGAS, IDIS, Santiago de Compostela, Spain, [15]Moores Cancer Center, University of California San Diego, La Jolla, CA, USA, [16]Oncology and Genetics Unit, Complejo Hospitalario Universitario de Vigo, CHUVI, SERGAS, Fundacion Biomedica Galicia Sur, Vigo, Spain, [17]Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France, [18]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, [19]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, [20]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [21]Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, [22]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg,

Germany, [23]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany, [24]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, [25]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany, [26]University of Tübingen, Tübingen, Germany, [27]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum, Bochum, Germany, [28]Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland, [29]Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland, [30]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland, [31]VIB Center for Cancer Biology, VIB, Leuven, Belgium, [32]Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Belgium, [33]Department of Radiation Oncology, University Hospitals Leuven, University of Leuven, Leuven, Belgium, [34]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, [35]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA, [36]Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh Medical School, Edinburgh, UK, [37]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [38]Academic Unit of Molecular Oncology, Department of Oncology and Metabolism, University of Sheffield, Sheffield UK, [39]Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK, [40]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK, [41]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK, [42]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands, [43]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, USA, [44]Department of Oncology, School of Medicine, Johns Hopkins University, USA, [45]Research Group Genetic Cancer Epidemiology, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

*Corresponding author.

†These authors contributed equally to this work.

ABSTRACT

Background: Polygenic risk scores (PRS) for breast cancer can be used to stratify the population into groups at substantially different levels of risk. Combining PRSs and environmental risk factors will improve risk prediction; however, integrating PRS into risk prediction models requires evaluation of their joint association with known environmental risk factors.

Methods: Analyses were based on data from 20 studies, datasets analyzed ranged from 3,453 to 23,104 invasive breast cancer cases and similar numbers of controls, depending on the analyzed environmental risk factor. We evaluated joint associations of a 77-single nucleotide polymorphism (SNP) PRS with reproductive history, alcohol consumption, menopausal hormone therapy (MHT), height and body mass index (BMI). We tested the null hypothesis of multiplicative joint associations for PRS and each of the environmental factors, and performed global and a tail-based goodness-of-fit tests in logistic regression models. The outcomes were breast cancer overall and by estrogen receptor (ER) status.

Results: The strongest evidence for a non-multiplicative interaction with the 77-SNP PRS was for alcohol consumption (P-interaction=0.009), adult height (P-interaction =0.025) and current use of combined MHT (P-interaction =0.038) in ER-positive disease. Risk associations for these factors by percentiles of PRS did not follow a clear dose-response. In addition, global and tail-based goodness of fit tests showed little evidence for departures from a multiplicative risk model, with alcohol consumption showing the strongest evidence for ER-positive disease (P=0.013 for global and 0.18 for tail-based test).

Conclusions: The combined effects of the 77-SNP PRS and environmental risk factors for breast cancer are generally well described by a multiplicative model. Larger studies are required to confirm possible departures from the multiplicative model for individual risk factors, and assess models specific for ER-negative disease.

Key words: breast cancer, genetic susceptibility, gene-environment interactions, risk prediction, epidemiology

Key Messages

- The combined effects of a polygenic risk score (PRS) derived from 77 single nucleotide polymorphisms (SNPs) and environmental risk factors for ER-positive breast cancer were generally well described by a multiplicative risk model.
- Analyses suggested non-multiplicative interactions of the 77-SNP PRS with alcohol consumption, height and menopausal hormone therapy (MHT) that did not follow a clear dose-response.
- Larger studies are required to confirm possible departures from the multiplicative model for individual risk factors, and assess models specific for ER-negative disease.

INTRODUCTION

Both inherited genetic factors and “environmental” factors, broadly defined as reproductive events (menarche, pregnancy, breast feeding and menopause), modifiable lifestyle (overweight/obesity, alcohol consumption, and physical activity); exogenous hormone medications (oral contraceptive pill and hormone replacement therapy) and medical history, play important roles in breast cancer etiology.[1] Genome-wide association studies have identified more common, low risk single nucleotide polymorphisms (SNPs) that in combination can substantially influence the risk of developing breast cancer.[2, 3] We previously described a 77-SNP polygenic risk score (PRS) for breast cancer; women in the top 1% of the PRS were at three-fold increased risk of developing the disease compared with women in the middle quintile.[4] This PRS explained ~12.6% of the familial relative risk (FRR) of breast cancer. The strength of the association (as measured by the relative risk per standard deviation) between the 77-SNP PRS and breast cancer risk decreased with increasing age. The association was similar in women with and without a family history, suggesting a multiplicative joint association of the PRS and other familial factors.[4]

In combination with environmental risk factors, the polygenic risk defined by the PRS and the residual FRR not explained by the PRS could result in substantial improvements in our ability to distinguish women at different levels of breast cancer risk in the general population, which could then be used to improve prevention and screening strategies for breast cancer.[5-8] Previous studies have indicated that established genetic and environmental risk factors are likely to combine multiplicatively in their associations with breast cancer risk.[9-12] A recent report evaluated interactions between a 24-SNP PRS and multiple environmental risk factors.[5] This study showed a good fit of a multiplicative risk model but had limited power to detect interactions, particularly at the extremes of the PRS. We have extended this study to evaluate the joint associations of the 77-SNP PRS and environmental risk factors for breast cancer using data from a larger multi-center study comprising 28,239 cases and 30,445 controls from 20 studies in the Breast Cancer Association Consortium (BCAC). Given that both environmental and genetic risk factors have been shown to differ by disease subtypes defined by estrogen receptor (ER) status,[13-15] analyses were performed for overall disease and separately for ER-positive and ER-negative disease. This study has immediate relevance as the 77 SNP PRS is currently being incorporated into risk prediction models for genetic counselling.

MATERIALS AND METHODS

Study sample

The study sample comprised 28,239 cases and 30,445 controls of European ancestry from 20 studies: two case-control studies nested in prospective cohorts, 8 population-based case-control and 10 non-population based case-control studies, all participating in the Breast Cancer Association Consortium (BCAC) (**Supplementary Tables 1 and 2**). Eligible studies had at least 200 cases and 200 controls with genotype data and information on at least one of the environmental risk factors of interest. Studies that oversampled cases with family history of breast cancer were excluded.

We excluded participants if they were male, were not of European descent (as defined by genome-wide genotype data), or had a missing value for age (age at diagnosis or interview for cases or controls, respectively). Statistical models included subjects with complete data on the specific environmental variable of interest and the adjustment variables. The number of participants available for analysis, therefore, varied by the investigated environmental factor. We also excluded prevalent cases from the cohort studies (date of diagnosis before baseline questionnaire) and cases from case-control studies interviewed more than five years after their diagnosis.

The relevant ethics committees approved individual studies and all study subjects gave written informed consent.

Data harmonization and variable definitions

Data from different studies were harmonized according to a common data dictionary. A quality assurance procedure was applied that included range and logic checks and comparisons of variable distributions within and between studies. Time-dependent variables were assessed at a reference date defined as the date of diagnosis for cases and the date of interview for controls in case-control studies. For cohort studies (MCCS and UKBGS), the reference date was the date of last follow-up questionnaire if data were available; otherwise date of baseline questionnaire was used as the reference.[9] The median time between the dates of last interview and diagnosis for cohort study participants was 2.0 years for UKBGS and 7.5 years for MCCS. Because we did not have data on menopausal status, we used the median age (54 years) as a surrogate: women aged <54 years were considered premenopausal and women aged ≥54 years postmenopausal.[9]

Seven risk factors for breast cancer were considered: age at menarche, ever being parous, age at first full-term pregnancy (AFTP), adult body mass index (BMI) in postmenopausal women, adult body height, current use of estrogen-progesterone menopausal hormone therapy (MHT), and lifetime average intake of alcohol. Current use of estrogen-progesterone MHT was defined as use within 6 months prior to the reference date. For case-control studies, BMI was calculated based on usual adult weight or weight one year prior to the reference date, if available (studies ABCFS, BREOGAN, CECILE, GENICA, MARIE, MCBCS, PBCS, SASBAC). If this variable was not available, body weight in early adulthood was used as a surrogate (studies ESTHER, pKARMA, SEARCH). Weight reported at the time of diagnosis or interview in case-control studies was not used to avoid disease effects on weight. For the two prospective cohort studies (MCCS, UKBGS), we used weight reported at the baseline interview (prior to diagnosis). Continuous variables (i.e. age at menarche, AFTP, alcohol, height and BMI) were modelled both as continuous and categorical variables; categories are shown in **Supplementary Table 3**.

Genotyping and Imputation

The rsnumbers for the 77 SNPs included in this report are shown in **Supplementary Table 4**. Genotype data for 76 of the 77 SNPs included in the PRS were generated as part of the Collaborative Oncological Gene-environment Study (COGS; www.nature.com/icogs) using an Illumina iSelect array (iCOGS) in all studies except BREOGAN. One SNP (rs78540526) was not genotyped but imputed using SHAPEIT and IMPUTEv2, using 5Mb non-overlapping intervals, as previously described.[16] Genotyping methods and quality control criteria have also been previously described.[17] Briefly, SNPs were excluded if the call rate was <95%, *P* for Hardy-Weinberg-Equilibrium test $<10^{-7}$, the concordance rate in duplicate samples was <98%, or if the SNP was monomorphic. Study participants were excluded from analyses if the overall genotyping call rate was <95% over the whole iCOGS array or if heterozygosity deviated from that expected in the general population (either lower or higher, $P < 10^{-6}$).

Genotyping for BREOGAN was performed at the Spanish National Genotyping Center (CeGen-ISCIII), using the Sequenom MassARRAY Genotyping system (technology iPLEX GOLD) following the manufacturer's instructions. The SNPs were analyzed using 4 assays (Assay Design v4 software) and genotyping calls were generated using the software Typer analyzer v4.0.20. The quality criteria described above were applied. The assay for rs7726159 failed and imputation of genotypes could not be conducted for this SNP or rs78540526 because of lack of other genotypes in BREOGAN. Therefore, only data on 75 SNPs were available for this study.

Statistical Methods

We investigated interactions between environmental risk factors for breast cancer and the PRS as a measure of the combined effects of 77 established SNPs on breast cancer risk. The calculation of the PRS for overall breast cancer and the PRS specific for ER-positive and ER-negative disease has been previously described.[4] Briefly, the PRS was derived for each study subject using the formula:

$$\text{PRS}=\beta_1 x_1+\beta_2 x_2+\ldots\beta_\kappa x_\kappa\ldots.+\beta_n x_n$$

where β_k was the per-allele log odds ratio (OR) for breast cancer associated with the minor allele for SNP *k*, x_k was the number of alleles for that same SNP (0, 1 or 2), and n=77 was the total number of SNPs (except for BREOGAN where we derived a 75 SNP PRS). To derive the ER-positive PRS, allele counts were weighted by ER-positive specific effect estimates; likewise, ER-negative specific effect estimates were used to derive the ER-negative PRS. The log ORs for each of the SNPs used to calculate the PRS were estimated using data in this report and are provided in **Supplementary Table 4**. These estimates are very close to those in our previous report,[4] which is expected given the large overlap in study populations.

ORs and 95% confidence intervals (CIs) were estimated using logistic regression models for overall breast cancer risk and by ER status of the tumor. Initial analyses included all studies with available data, regardless of study design, and considered each environmental variable one at a time. Models were adjusted for study (indicator variables), age and seven ancestry-informative principal components (for models including PRS). All models also included an interaction term between study design (population-based/cohort vs non-population based; see **Supplementary Table 1**) and the environmental variable of interest, to account for potential heterogeneity of main effects by design. Because estimates of main effects of environmental variables from non-population-based designs are prone to bias, we only reported results from population-based/cohort studies. However, interaction estimates and statistical tests of interaction (see below) are based on data from all studies. In models including current use of combined (estrogen-progesterone) MHT, users of combined MHT were compared with never users of any MHT and were further adjusted for use of MHT preparations other than combined therapy. MHT analyses were restricted to postmenopausal women. To assess interaction, we used a likelihood ratio test (LRT) comparing models with and without interaction terms for the PRS as a continuous variable and each of the environmental variables (modelled as continuous variables when appropriate).[12] Separate models were fit for each PRS and environmental risk factor combination.

To assess the goodness of fit of a multiplicative model, we also performed, for each risk factor, a global goodness of fit test and a recently developed tail-based goodness of fit test to assess deviations from logistic models at the extremes of the risk distribution.[18] For goodness of fit tests, analyses were restricted to population-based/cohort studies to remove the contribution of non-population based studies to the main effect estimates of environmental risk factors as these are more prone to biases. The goodness of fit tests were not fit for ER-negative disease, as the number of controls and the number of cases available for analysis was too small to provide reliable estimates, particularly in the tails.

The statistical analysis was conducted using SAS 9.3 and R (version 3.0.2). All tests performed were two-sided.

RESULTS

A total of 28,241 cases and 30,445 controls from 20 studies contributed data to at least one analysis. The numbers of cases and controls from each of the studies are shown in **Supplementary Table 2.** The associations between the 77-SNP PRS for overall and subtype specific breast cancer are shown in **Supplementary Figure 1**. As shown previously using a similar study population as in this report,[4] associations were stronger for ER-positive than ER-negative disease.

Associations of environmental risk factors in relation to overall and ER-positive breast cancer risk, based on data from population-based or cohort studies were of the expected magnitude and direction (**Supplementary Table 3**). Associations for nulliparity and MHT use differed by ER status of the tumor (P_{het}<0.003) and none of the environmental risk factors showed test for associations with ER-negative disease with P<0.05. Because of the relatively small number of ER-negative cases, we focused the presentation of interaction analyses on all breast cancers or ER-positive breast cancer.

Results from our primary analyses of interaction between PRS and individual environmental risk factors are shown in **Table 1**. The strongest evidence for non-multiplicative joint associations in ER-positive disease, as assessed by a trend in the OR by PRS level, was for alcohol consumption (LRT P = 0.009 based on 3,453 cases and 3,708 controls with available data), adult height (LRT P=0.025 based on 20,417 cases ad 18,412 controls) and current use of MHT (LRT P=0.038 based on 5,201 cases and 5,697 controls; **Table 1**). These interaction analyses were based on a study sample ranging from 3,453 cases and 3,708 controls for average lifetime intake of alcohol, to 23,104 cases and 25,914 controls for parity, and multiplicative interaction parameters showed no evidence for heterogeneity between population-based/cohort and non-population-based study designs (**Supplementary Table 5)**. We found no evidence for interactions in ER-negative disease (**Table 1**). **Figure 1** shows the estimated ORs (95%CI) for the risk of ER-positive breast cancer and each of the environmental risk factors stratified by percentiles of the PRS (see **Supplementary Figure 2** for results for overall breast cancer and by ER status). It should be noted that interaction tests in **Table 1** considered PRS as a continuous variable rather than in percentile categories as shown in the Figures. Estimated ORs by PRS percentiles for the three environmental factors in **Table 1** did not show clear dose-response relationships, particularly for alcohol consumption and adult height (**Figure 1**): the interaction for alcohol was mainly driven by the relatively large OR estimate for the lowest percentile of the PRS; the OR estimates for height were stronger for the middle categories of PRS; and the ORs for MHT showed more of a dose-response pattern, although not entirely consistent across categories of PRS.

Global and tail-based goodness of fit tests for models including the 77-SNP PRS and each of the environmental factors were performed in population-based or cohort studies only. These analyses did not show substantial evidence for departures from the multiplicative model, except alcohol consumption in ER-positive disease (P=0.013 for global and 0.18 for tail based tests; **Table 2**).

DISCUSSION

Our analyses indicate that the combined effects of the 77-SNP PRS and environmental risk factors (reproductive history, MHT use, adult height, BMI and alcohol intake) for breast cancer are generally consistent with a multiplicative model on the relative risk scale. An important consequence of the multiplicative model is that the absolute risk associated with each environmental factor would be larger among women at high genetic risk; this could be relevant to counselling and intervention studies. The observed evidence for non-multiplicative joint associations of PRS and alcohol intake, height and MHT use requires confirmation in larger studies.

Previous reports have shown that most SNPs and environmental risk factors, considered pairwise, combine multiplicatively.[9-12, 19] It is plausible, however, that groups of susceptibility variants could in combination interact with environmental risk factors. We therefore evaluated the joint association with a PRS summarizing the risk conferred from 77 SNPs (a straightforward and efficient approach, since there is little evidence for non-multiplicative interactions among SNPs).[4] This is relevant since models combining multiple SNPs in the form of PRSs are being used in risk prediction models that integrate genetic and environmental factors.[5, 8, 20, 21] A recent report evaluated interactions between a 24-SNP PRS and environmental risk factors (age at first birth, parity, age at menarche, height, menopausal status, age at menopause, BMI, MHT use, alcohol consumption and smoking status) based on analyses of data from 17,171 cases and 19,862 controls sampled from eight prospective cohort studies in the Breast and Prostate Cancer Cohort Consortium (BPC3).[5] This study found no evidence for departures from the multiplicative model for any of the risk factors evaluated, which is generally consistent with the goodness-of-fit test performed in population-based studies in this report. The BPC3 findings do not support the observed interactions between the 77-SNP PRS and alcohol consumption, height and MHT use in our report. Although it is possible that interactions are evident with the extended 77-SNP PRS but not the 24-SNP PRS used in BPC3, they need to be replicated in independent studies with appropriate study designs, particularly in view of the lack of a clear dose-response pattern for the interactions in our report. Our result should also be interpreted with caution because of multiple hypothesis testing and the relatively low power (as reflected by the wide confidence intervals in estimates of interaction parameters) that can lead to a higher probably of false positive findings for a given significance level.[22]

The 77 SNP PRS in our analysis is more predictive than the 24 SNP PRS evaluated in the BPC3 report since it includes all 24 SNPs plus additional SNPs identified in subsequent genome-wide association studies. However, the 77-SNP PRS could be over-fitted since our study population largely overlaps with populations in genome wide association studies that lead to the discovery of most of known SNPs.[17, 23] Nevertheless, over-fitting of the PRS is unlikely to bias the assessment of interactions with environmental risk factors.

A strength of our study is the large total sample size; however, data for some risk factors, particularly alcohol consumption and use of MHT, was only available from a subset of studies or was missing for a substantial number of participants. In addition, our report includes studies with different study designs: ten of 20 studies were non-population-based case-control studies that are prone to biases in assessing associations with environmental risk factors. To address this limitation, we included an interaction term for the environmental exposure and study design (population-based (including cohorts) versus non-population-based), and used only main effects estimates from population-based studies. In contrast, we used all data available for estimation of multiplicative interaction parameters since they are less susceptible to differential measurement error in case-control studies than main effect parameters,[24] and showed no evidence for heterogeneity across study designs.

Interactions with environmental risk factors, such as benign breast disease, mammographic breast density, oral contraceptive use or physical activity, are possible but could not be evaluated in this report due to sparse or lack of available data. A recent report based on a 76-SNP PRS and Breast Imaging Reporting and Data System (BI-RADS) breast density did not show evidence for non-multiplicative joint associations, albeit in a relatively small study including 1,643 cases and 2,397 controls.[21] Larger studies are needed to further evaluate the joint associations between PRS and these factors. More data than that included in this report will also be required to assess the joint effects for ER-negative disease, where the sample sizes and effect sizes for some factors are smaller.

In summary, our results provide support for the assumption of multiplicative joint associations between PRS and environmental risk factors in the development of risk prediction models for breast cancer; however, small departures are possible and require further investigation. Risk prediction tools based on validated models that can be easily implemented in clinical practice will be needed for the evaluation and ultimate adoption of risk-stratification-based strategies in breast cancer prevention and screening.

TABLES AND FIGURES

Table 1. Odds ratios and 95% confidence intervals for multiplicative interaction between polygenic risk score and environmental risk factors of breast cancer, for all and ER-positive breast cancers, based on population-based and non-population-based studies.

Table 2. Goodness of fit test p-values for overall breast cancer and estrogen receptor positive breast cancer, based on population-based studies.

Figure 1. Odds ratios and 95% confidence intervals for breast cancer risk factors by percentiles of the 77-SNP polygenic risk score (PRS) specific for ER-positive breast cancer, based on population-based and non-population-based studies. FFTP: First full-term pregnancy.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Description of BCAC studies included in the analysis of multiplicative interaction between environmental risk factors and 77-SNP polygenic risk score (PRS).

Supplementary Table 2. List of participating studies and number of subjects of European descent included in at least one GxE analysis.

Supplementary Table 3. Associations of environmental risk factors with breast cancer risk, overall and by ER status of the tumor, based on population-based studies.

Supplementary Table 4. SNPs included in polygenic risk score and effect sizes for association with breast cancer or subtypes of the disease.

Supplementary Table 5. Odds ratios and 95% confidence intervals for multiplicative interaction between 77-SNP polygenic risk score (PRS) and environmental risk factors of breast cancer by study design category.

Supplementary Figure 1. Odds ratios and 95% confidence intervals for percentiles of the 77-SNP polygenic risk score (PRS), for all, ER-positive breast cancer and ER-negative breast cancer, based on population-based and non-population-based studies.

Supplementary Figure 2. Odds ratios and 95% confidence intervals for breast cancer risk factors by percentiles of the 77-SNP polygenic risk score (PRS) for all, ER-positive breast cancer and ER-negative breast cancer, based on population-based and non-population-based studies.

FUNDING

This work was supported by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). The Australian Breast Cancer Family Study (**ABCFS**) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. The **ABCS** study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]. The work of the **BBCC** was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BREast Oncology GAlician Network (**BREOGAN**) is funded by FIS ISCIII/PI12/02125 Acción Estratégica de Salud del Instituto de Salud Carlos III, FEDER; FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, CHUVI Vigo Hospital, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I+D e I+D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain.We thank José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS. The **CECILE** study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The **CGPS** was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The **ESTHER** study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). The **GENICA** was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. The **KBCP** was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, and by the strategic funding of the University of Eastern Finland. **LMBC** is supported by the 'Stichting tegen Kanker'. Diether Lambrechts is supported by the FWO. The **MARIE** study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. The **MCBCS** was supported by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. The work of **MTLGEBCS** was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the “CIHR Team in Familial Risks of Breast Cancer” program – grant # CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade – grant # PSR-SIIRI-701. **MCCS** cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. The **PBCS** was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The **KARMA** study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. The **SASBAC** study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. The **SBCS** was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now. **SEARCH** is funded by a programme grant from Cancer Research UK [C490/A10124] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The **UKBGS** is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre.

ACKNOWLEDGEMENTS

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. This study would not have been possible without the contributions of the following:

ABCFS: Maggie Angelakos, Judi Maskiell, Gillian Dite; **ABCS**: Blood bank Sanquin, The Netherlands; **BREOGAN**: This study would not have been possible without the contributions of the following: Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BREOGAN), José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS; Joaquín González-Carreró and the staff of the Department of Pathology and Biobank of University Hospital

Complex of Vigo, Instituto de Investigacion Biomedica Galicia Sur, SERGAS, Vigo, Spain; **CGPS**: Staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. **ESTHER**: Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, Katja Butterbach, Katarina Cuk, Kai-Uwe Saum; **GENICA**: The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [HB, Wing-Yee Lo, Christina Justenhoven], German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) [HB], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany [Yon-Dschun Ko, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [Ute Hamann], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [TB, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]; **KBCP**: Eija Myöhänen, Helena Kemiläinen; **LMBC**: Gilian Peuteman, Thomas Van Brussel, Evy Vanderheyden and Kathleen Corthouts; **MARIE**: Petra Seibold, Dieter Flesch-Janys, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels; **MTLGEBCS**: We would like to thank Martine Tranchant (CHU de Québec Research Center), Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management and skillful technical assistance. J.S. is Chairholder of the Canada Research Chair in Oncogenetics; **PBCS**: Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner; **pKARMA**: The Swedish Medical Research Counsel; **SASBAC**: The Swedish Medical Research Counsel; **SBCS**: Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian and Malcolm W.R. Reed; **SEARCH**: The SEARCH and EPIC teams; **UKBGS**: We thank Breast Cancer Now and the

Institute of Cancer Research for support and funding of the Breakthrough Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre.

REFERENCES

1. Colditz GA, Baer HJ, Tamimi RM. Breast Cancer. In: Schottenfeld D, Fraumeni JF, editors. *Cancer epidemiology and prevention*. Oxford; New York: Oxford University Press; 2006.

2. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nature genetics* 2002; **31**: 33-6.

3. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *New England Journal of Medicine* 2008; **358**: 2796-803.

4. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute* 2015; **107**.

5. 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 Oncol* 2016.

6. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature reviews Genetics* 2016; **17**: 392-406.

7. 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* 2013; **45**: 349-51.

8. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *Journal of the National Cancer Institute* 2014; **106**.

9. Nickels S, Truong T, Hein R, et al. Evidence of Gene-Environment Interactions between Common Breast Cancer Susceptibility Loci and Established Environmental Risk Factors. *PLoS Genet* 2013; **9**: e1003284.

10. Travis RC, Reeves GK, Green J, et al. Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. *Lancet* 2010; **375**: 2143-51.

11. Barrdahl M, Canzian F, Joshi AD, et al. Post-GWAS gene-environment interplay in breast cancer: results from the Breast and Prostate Cancer Cohort Consortium and a meta-analysis on 79,000 women. *Human molecular genetics* 2014; **23**: 5260-70.

12. Rudolph A, Milne RL, Truong T, et al. Investigation of gene-environment interactions between 47 newly identified breast cancer susceptibility loci and environmental risk factors. *International journal of cancer Journal international du cancer* 2015; **136**: E685-96.

13. Yang XR, Chang-Claude J, Goode EL, et al. Associations of Breast Cancer Risk Factors With Tumor Subtypes: A Pooled Analysis From the Breast Cancer Association Consortium Studies. *Journal of the National Cancer Institute* 2011; **103**: 250-63.

14. Broeks A, Schmidt MK, Sherman ME, et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Human molecular genetics* 2011: ddr228.

15. Garcia-Closas M, Chanock S. Genetic susceptibility loci for breast cancer by estrogen receptor status. *Clin Cancer Res* 2008; **14**: 8000-9.

16. 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. *Nature genetics* 2015; **47**: 373-80.

17. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature genetics* 2013; **45**: 353-61.

18. Song M, Kraft P, Joshi AD, Barrdahl M, Chatterjee N. Testing calibration of risk models at extremes of disease risk. *Biostatistics* 2015; **16**: 143-54.

19. Campa D, Kaaks R, Le Marchand L, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *Journal of the National Cancer Institute* 2011; **103**: 1252-63.

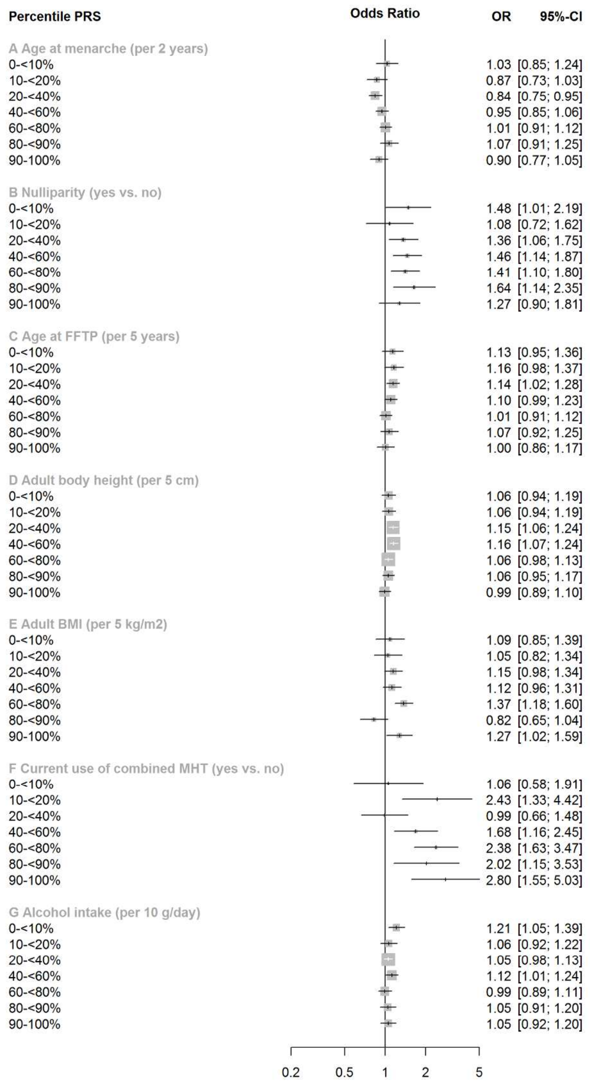
20. Shieh Y, Hu D, Ma L, et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast cancer research and treatment* 2016; **159**: 513-25.

21. Vachon CM, Pankratz VS, Scott CG, et al. The contributions of breast density and common genetic variation to breast cancer risk. *Journal of the National Cancer Institute* 2015; **107**.

22. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute* 2004; **96**: 434-42.

23. Garcia-Closas M, Couch FJ, Lindstrom S, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nature genetics* 2013; **45**: 392-8.

24. Garcia-Closas M, Thompson WD, Robins JM. Differential misclassification and the assessment of gene-environment interactions in case-control studies. *American journal of epidemiology* 1998; **147**: 426-33.



Odds ratios and 95% confidence intervals for breast cancer risk factors by percentiles of the 77-SNP polygenic risk score (PRS) specific for ER-positive breast cancer, based on population-based and non-population-based studies. FFTP: First full-term pregnancy.

190x381mm (200 x 200 DPI)

Table 1. Odds ratios and 95% confidence intervals for multiplicative interaction between polygenic risk score and environmental risk factors of breast cancer, for all, ER-positive breast cancer and ER-negative breast cancer, based on population-based and non-population-based studies

Environmental Factor	**N Studies**	**N cases / controls**	**$\mathrm{OR}_{\mathrm{int}}$ (95% CI)[1]**	***Pint***	**N cases / control**	**$\mathrm{OR}_{\mathrm{int}}$ (95% CI)[1]**	***Pint***	**N cases / control**	**$\mathrm{OR}_{\mathrm{int}}$ (95% CI)[1]**	***Pint***
		All breast cancers			**ER positive breast cancer**			**ER negative breast cancer**		
Age at menarche (per 2 years)	17	18175 / 20366	1.02 (0.96 - 1.08)	0.50	12664 / 20366	1.02 (0.96 - 1.08)	0.62	2995 / 20366	1.00 (0.88 - 1.14)	0.98
Nulliparity (yes vs. no)	19	23104 / 25914	1.05 (0.93 - 1.19)	0.45	16293 / 25914	1.04 (0.92 - 1.18)	0.55	3719 / 25914	1.11 (0.84 - 1.45)	0.48
Age at first full-term pregnancy (per 5 years)	16	15523 / 17623	0.96 (0.91 - 1.01)	0.10	10807 / 17623	0.96 (0.91 - 1.01)	0.15	2557 / 17623	0.92 (0.81 - 1.03)	0.14
Alcohol consumption (per 10g/day)	5	3453 / 3708	0.90 (0.82 - 0.98)	0.016	2661 / 3708	0.89 (0.82 - 0.97)	0.009	538 / 3708	1.16 (0.92 - 1.47)	0.22
Adult height (per 5 cm)	18	20417 / 18412	0.96 (0.92 - 0.99)	0.012	14525 / 18412	0.96 (0.92 - 0.99)	0.025	3389 / 18412	0.97 (0.90 - 1.04)	0.41
Adult BMI (per 5 kg/m^2)	12	8188 / 6717	0.96 (0.88 - 1.05)	0.45	6007 / 6717	0.97 (0.89 - 1.06)	0.48	1229 / 6717	0.92 (0.77 - 1.10)	0.35
Current use of combined MHT (yes vs. never)2	7	5201 / 5697	1.27 (0.95 - 1.70)	0.10	4147 / 5697	1.34 (1.02 - 1.77)	0.038	763 / 5697	0.95 (0.50 - 1.79)	0.87

[1] Adjusted for reference age, study, ancestry-informative principal components and an interaction term between environmental factor and study design (population-based vs. non-population-based). Models used to assess association with use of combined MHT have been further adjusted use of other MHT preparations.

[2] Postmenopausal women only

ER: estrogen receptor; $\mathrm{OR}_{\mathrm{int}}$: odds ratio for interaction; CI: confidence interval

Table 2. Goodness of fit test p-values for overall breast cancer and estrogen receptor positive breast cancer, based on population-based studies.

	Overall breast cancers				ER positive breast cancer			
Variables included in models	**N Studies**	**N cases / controls**	**Tail-based goodness-of-fit test**	**Global goodness-of-fit test**	**N Studies**	**N cases / controls**	**Tail-based goodness-of-fit test**	**Global goodness-of-fit test**
Single risk factor models with 77-SNP PRS								
Age at menarche	10	6209 / 6207	0.758	0.776	10	4320 / 6207	0.869	0.563
Nulliparity	10	6507 / 6578	0.639	0.888	10	4517 / 6578	0.540	0.085
Age at first full-term pregnancy [2]	9	5060 / 5317	0.760	0.562	9	3505 / 5317	0.445	0.306
Alcohol consumption	5	3453 / 3708	0.763	0.565	5	2661 / 3708	0.175	0.013
Adult body height	10	6462 / 6522	0.923	0.875	10	4476 / 6522	0.917	0.219
Adult BMI	8	2958 / 3343	0.956	0.933	8	2099 / 3343	0.563	0.352
MHT [3]	11	5060 / 5208	0.773	0.606	11	3636 / 5208	0.354	0.489
Multiple risk factor models with 77- SNP PRS								
Adult BMI + MHT + BMI*MHT [3]	5	2065 / 2417	0.205	0.655	5	1556 / 2417	0.386	0.494
All environmental factors with BMI*MHT + age + family history	3	1012 / 1161	0.179	0.251	3	847 / 1161	0.679	0.476

[1] always adjusted for study

[2] in parous women only

[3] Menopausal hormone therapy (MHT) categorized as follows: category 1: premenopausal women, irrespective of MHT use; category 2: postmenopausal women who never used MHT; category 3: postmenopausal women who used any kind of MHT in the time period up to six month before reference age; category 4: postmenopausal women who used estrogen-progestogen therapy (EPT) in the last six month before reference age; category 5: postmenopausal women who used any other kind of MHT despite EPT in the last six month before reference age

Age, age at menarche, age at first full time pregnacy, alcohol, height, BMI are in categories