Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States

Paige Maas, PhD, Myrto Barrdahl, PhD, Amit D. Joshi, PhD, Paul L. Auer, PhD, Mia M. Gaudet, PhD, Roger L. Milne, PhD, Fredrick R. Schumacher, PhD, William F. Anderson, MD, MPH, David Check, BS, Subham Chattopadhyay, BS, Laura Baglietto, PhD, Christine D. Berg, PhD, Stephen J. Chanock, MD, David G. Cox, PhD, Jonine D. Figueroa, PhD, Mitchell H. Gail, MD, PhD, Barry I. Graubard, PhD, Christopher A. Haiman, ScD, Susan E. Hankinson, ScD, Robert N. Hoover, MD, ScD, Claudine Isaacs, MD, Laurence N. Kolonel, PhD, Loic Le Marchand, MD, PhD, I-Min Lee, ScD, Sara Lindström, PhD, Kim Overvad, PhD, Isabelle Romieu, MD, MPH, ScD, Maria-Jose Sanchez, PhD, Melissa C. Southey, PhD, Daniel O. Stram, PhD, Rosario Tumino, PhD, Tyler J. VanderWeele, PhD, Walter C. Willett, MD, DrPh, Shumin Zhang, PhD, Julie E. Buring, ScD, Federico Canzian, PhD, Susan M. Gapstur, PhD, Brian E. Henderson, MD, David J. Hunter, MBBS, ScD, Graham G Giles, PhD, Ross L. Prentice, PhD, Regina G. Ziegler, PhD, MPH, Peter Kraft, PhD, Montse Garcia-Closas, MPH, DrPH, and **Nilanjan Chatterjee, PhD**

Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland (Maas, Anderson, Check, Chattopadhyay, Chanock, Figueroa, Gail, Graubard, Hoover, Ziegler, Garcia-Closas, Chatterjee); Division of Cancer Epidemiology, German

---

Corresponding Author: Nilanjan Chatterjee, PhD, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe St, Baltimore, MD 21205 (nilanjan@jhu.edu).

Cancer Research Center (DKFZ), Heidelberg, Germany (Barrdahl); Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (Joshi, Lindström, Hunter, Kraft); Fred Hutchinson Cancer Research Center, Seattle, Washington (Auer, Prentice); School of Public Health, University of Wisconsin-Milwaukee, Milwaukee (Auer); Epidemiology Research Program, American Cancer Society, Atlanta Georgia (Gaudet, Gapstur); Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia (Milne, Baglietto, Giles); Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia (Milne, Baglietto, Giles); Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles (Schumacher, Haiman, Stram, Henderson); Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland (Berg); INSERM U1052 – Cancer Research Center of Lyon, Centre Léon Bérard, Lyon, France (Cox); Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, England (Cox); Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst (Hankinson); Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts (Hankinson); Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC (Isaacs); Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu (Kolonel); University of Hawaii Cancer Center, Honolulu (Le Marchand); Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts (Lee, Zhang, Buring); Department of Public Health, Aarhus University, Aarhus, Denmark (Overvad); Nutrition and Metabolism Section, International Agency for Research on Cancer, Lyon, France (Romieu); Escuela Andaluza de Salud Pública. Instituto de Investigación Biosanitaria ibs.GRANADA. Hospitales Universitarios de Granada/Universidad de Granada, Granada, Spain (Sanchez); CIBER de Epidemiología y Salud Pública (CIBERESP), Spain (Sanchez); Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Australia (Southey); Cancer Registry and Histopathology Unit, "Civic-M.P.Arezzo" Hospital, ASP Ragusa, Italy (Tumino); Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (VanderWeele); Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (VanderWeele); Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (Willett); German Cancer Research Center (DKFZ), Heidelberg, Germany (Canzian); Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia (Giles); University of Washington, School of Public Health and Community Medicine, Seattle (Prentice); Breakthrough Breast Cancer Research Centre, Division of Genetics and Epidemiology, The Institute of Cancer Research, London, England (Garcia-Closas); Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Chatterjee); Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, Maryland (Chatterjee)

Abstract

IMPORTANCE—An improved model for risk stratification can be useful for guiding public health strategies of breast cancer prevention.

OBJECTIVE—To evaluate combined risk stratification utility of common low penetrant single nucleotide polymorphisms (SNPs) and epidemiologic risk factors.

DESIGN, SETTING, AND PARTICIPANTS—Using a total of 17 171 cases and 19 862 controls sampled from the Breast and Prostate Cancer Cohort Consortium (BPC3) and 5879 women participating in the 2010 National Health Interview Survey, a model for predicting absolute risk of breast cancer was developed combining information on individual level data on epidemiologic risk factors and 24 genotyped SNPs from prospective cohort studies, published estimate of odds ratios for 68 additional SNPs, population incidence rate from the National Cancer Institute-Surveillance, Epidemiology, and End Results Program cancer registry and data on risk factor distribution from nationally representative health survey. The model is used to project the distribution of absolute risk for the population of white women in the United States after adjustment for competing cause of mortality.

EXPOSURES—Single nucleotide polymorphisms, family history, anthropometric factors, menstrual and/or reproductive factors, and lifestyle factors.

MAIN OUTCOMES AND MEASURES—Degree of stratification of absolute risk owing to nonmodifiable (SNPs, family history, height, and some components of menstrual and/or reproductive history) and modifiable factors (body mass index [BMI; calculated as weight in kilograms divided by height in meters squared], menopausal hormone therapy [MHT], alcohol, and smoking).

RESULTS—The average absolute risk for a 30-year-old white woman in the United States developing invasive breast cancer by age 80 years is 11.3%. A model that includes all risk factors provided a range of average absolute risk from 4.4% to 23.5% for women in the bottom and top deciles of the risk distribution, respectively. For women who were at the lowest and highest deciles of nonmodifiable risks, the 5th and 95th percentile range of the risk distribution associated with 4 modifiable factors was 2.9% to 5.0% and 15.5% to 25.0%, respectively. For women in the highest decile of risk owing to nonmodifiable factors, those who had low BMI, did not drink or smoke, and did not use MHT had risks comparable to an average woman in the general population.

CONCLUSIONS AND RELEVANCE—This model for absolute risk of breast cancer including SNPs can provide stratification for the population of white women in the United States. The model can also identify subsets of the population at an elevated risk that would benefit most from risk-reduction strategies based on altering modifiable factors. The effectiveness of this model for individual risk communication needs further investigation.

Breast cancer remains the most common form of cancer diagnosed in women in developed countries of the Western world, with an estimated 232 670 new cases diagnosed in 2014 in the United States alone.[1] The incidence of breast cancer is also reported to be rapidly rising in a number of developing countries, possibly owing to the congruence of a number of factors, including changes in lifestyle, behavioral patterns, and improved diagnostics, all results of economic growth.[2,3] Decades of epidemiologic research have led to the identification of a number of lifestyle and environmental breast cancer risk factors, including menstrual and/or reproductive history, use of hormones, anthropometry, and alcohol consumption, each typically explaining a modest proportion of the variation in disease risk.[4,5] However, when combined, the known risk factors could have a substantial effect on

breast cancer risk. More recently, genome-wide association studies (GWAS) have led to the identification of 92 common susceptibility loci marked by single nucleotide polymorphisms (SNPs).[6–8] These SNPs are each associated with only a small effect size but cumulatively explain substantial variation in risk.[9,10] The proportion of variation in risk explained by common genetic variation is likely to increase in the near future, after the completion of the OncoArray project[11] that is anticipated to detect many additional risk-associated variants for breast cancer.

As GWAS are rapidly expanding the spectrum of genetic risk factors for breast cancer, it is timely to evaluate how such information can be used to understand the distribution of breast cancer risk across populations and focus strategies for cancer prevention.[10,12,13] Following the discoveries from early GWAS, several studies[14–17] have reported only modest utility of SNPs for improving the discriminatory accuracy of breast cancer risk prediction models. However, a recent study[9] following numerous discoveries from the large Collaborative Oncological Geneenvironment Study (COGS) project indicated that a polygenic risk score (PRS) defined by the combination of 77 SNPs could be useful for providing substantial risk stratification of the population. As SNPs and certain other risk factors are nonmodifiable (ie, risk factors that cannot be modified or are unlikely to be modified with the purpose of altering breast cancer risk), it is unclear whether and how information on these nonmodifiable risk factors can guide primary cancer prevention efforts that intervene on modifiable risk factors. In a recent commentary,[10] we used a synthetic model, based on published estimates of risk parameters and the assumption of multiplicative gene-environment interaction, to show that a PRS defined by known SNPs can provide risk stratification to a degree that may be useful for prevention. For instance, it could be helpful in assessing individualized riskbenefit tradeoffs associated with the use of menopausal hormone therapy (MHT) and endocrine-based prevention strategies.

The goal of this study was to use data from prospective cohort studies participating in the Breast and Prostate Cancer Cohort Consortium (BPC3)[18,19] to develop a more empirical model for predicting absolute risk of invasive breast cancer. This model was then used to project the distributions of risk for the general population of white women in the United States, decomposed into modifiable and nonmodifiable risk components. We provide estimates of the number of breast cancers that would be preventable through risk factor modification in strata of the population at different levels of risk from nonmodifiable factors. Results from these projections provide new insight into the challenges and opportunities for risk-based targeted primary cancer prevention efforts.

Methods

Study Population

The BPC3 has previously been described in detail.[19–21] In short, it consists of 8 large, prospective cohorts from Europe, Australia, and the United States with genetic data and questionnaire information. The diagnosis of cases of breast cancer was confirmed by medical records and/or tumor registries. Analyses presented in this manuscript include only invasive breast cancer cases. We analyzed data available from the nested case-control samples within the cohorts selected for genetic studies. In these studies, subjects were

considered eligible controls if they were free of breast cancer until the diagnosis of breast cancer in the matched case subject. Matching criteria varied among cohorts, but age and menopausal status at baseline were used for all. The BPC3 project was approved by the ethical committee of the International Agency for Research on Cancer (IARC) for the EPIC cohort, by the Emory University Institutional Review Board for CPS-II cohort, by the Institutional Review Board of the University of Hawaii and University of Southern California for the MEC cohort, by the ethical committee of the Brigham and Women's Hospital for the NHS cohort and the NCI Institutional Review Board for the PLCO cohort.

The BPC3 study and published estimates of SNP odds ratios (ORs) were used to develop a logistic regression model that included a polygenic risk score (PRS), nonmodifiable risk factors other than the PRS (ie, family history, age at first birth, parity, age at menarche, height, menopausal status, and age at menopause), along with modifiable risk factors (ie, body mass index [BMI; calculated as weight in kilograms divided by height in meters squared], MHT use, level of alcohol consumption, and smoking status). The eMethods in the Supplement describe in detail all steps in the development of this model, which includes 92 known susceptibility SNPs (eTable 2 in the Supplement) and the other risk factors. Data on 24 SNPs genotyped in subjects in the BPC3 was initially used to derive a polygenic risk score for the 24 SNPs (PRS-24) by assuming additive associations on the log scale of the SNPs in the logistic regression model after adjustment for study, age at study entry, and family history. Data on the 24 genotyped SNPs was used to evaluate multiplicative interactions between individual SNPs and PRS-24 with other risk factors. We also used a recently developed tail-based χ^2 goodness of fit test[22] to assess possible deviations of risks estimated from a multiplicative model from true risks at the extremes of the risk distribution. Assuming the validity of the multiplicative model, we then derived a model based on all 92 known breast cancer SNPs (PRS-92) based on published ORs for the 68 remaining SNPs that were not genotyped in BPC3.

Absolute Risk Modeling

We built a model for absolute risk of invasive breast cancer for the population of white women in the United States by combining estimates of OR parameters obtained from the BPC3 and external GWAS studies, age-specific breast cancer rates from the US National Cancer Institute-Surveillance, Epidemiology, and End Results Program (NCI-SEER) and data on competing hazards for mortality available from the Center for Disease Control (CDC) WONDER database[23] (eMethods in the Supplement).

Projection of Absolute Risk Distribution for the Population of White Women in the United States

We projected the distribution of absolute risk for the population of white women in the United States based on the distribution of risk factors observed in nationally representative survey data from the National Health Interview Survey and National Health and Nutrition Examination Survey.[24–28] We assumed that risk factors and PRS-92 are independently distributed, conditional on family history. We then generated the distribution of PRS based on normal distribution theory (eMethods in the Supplement).

We further assessed the distribution of risk owing to modifiable risk factors (BMI, MHT use, alcohol, smoking) in categories defined by risk from nonmodifiable factors, including PRS-92. We estimated the proportion of breast cancer that could be prevented by shifting the whole population to the lowest level of modifiable risk within each strata of the population as defined by the nonmodifiable risk factors (eMethods in the Supplement).

Results

The analysis involved a total number of 17 171 cases and 19 862 controls from 8 prospective cohort studies, but the number of cases and controls with complete information in each study varied by risk factor (eTable 1 in the Supplement).

Assessment of Interactions and Risk Model Building

The additive model on the logistic scale for the SNP-risk associations in the PRS-24 risk model was adequate, even at the extremes of risk. Consistently, estimates of the ORs associated with deciles of a fitted logistic regression model for PRS-24 and family history closely followed their values predicted from the normal distribution theory for PRS (eMethods and eTable 3 in the Supplement).

Odds ratio estimates for individual risk factors from the fitted multivariate logistic regression model are shown in eFigure 1 in the Supplement. The association between risk and quantitative factors (height, number of children, age at first birth, and alcohol use) appeared to be nonlinear on the logistic scale; thus, in subsequent analysis, we modeled quantitative factors as categorical variables, defined by the deciles of their distributions in controls (eMethods and eTable 4 in the Supplement). Higher BMI was associated with increased risk only for postmenopausal women, and the strength of the association was stronger for patients who did not use MHT (eFigure 1 in the Supplement). We did not detect any statistically significant interactions between PRS-24 and individual risk factors in the categorical or the continuous modeling approaches (data not shown). We also performed an overall χ^2 goodness-of-fit test for this model using a tail-based method[22] and found that the model including both PRS-24 and all other risk factors in a multiplicative fashion (or additive in the logistic scale) fit the BPC3 data adequately.

The final risk model included main effects of the PRS-92 (genotyped PRS-24 plus simulated PRS-68, as described in our Methods section), main effects of all of the risk factors coded as categorical variables, and interaction terms involving menopausal status, BMI, and MHT variables (eMethods in the Supplement). The area under the receiver operating curve (AUC) for models with only questionnaire-based risk factors, only PRS-92, and both types of risk factors were 0.588,0.623, and 0.648, respectively (eFigure 2 in the Supplement).

Stratification of Absolute Breast Cancer Risk

Although AUC values were low to modest, the models, particularly the models including the PRS, led to substantial spread in the distribution of absolute risk for the population. For example, the absolute cumulative risk of a 30-year-old white woman in the United States developing invasive breast cancer over the next 50 years is 11.3% on average. A model based on PRS-92 and questionnaire-based risk factors could identify 5% of the population

with risk below 4.5% or above 22.0% (Figure 1). As risk accumulated over age, the degree of stratification of absolute risk provided by all the risk factors combined also increased with age (Figure 2). The percentage of the population that could be identified to be of moderate risk (twofold to 3-fold risk compared with the population average) and high risk (>3-fold risk compared with the population average) varied substantially among models (Table 1), with the most pronounced discrimination for the full model compared with models with only PRS-based or questionnaire-based risk factors.

Distribution of Modifiable and Nonmodifiable Breast Cancer Risk

The spread in the distribution of risk by the 4 modifiable risk factors (ie, BMI, MHT use, alcohol use, smoking) was larger for those substrata of the population that were at higher risk owing to nonmodifiable risk factors (Figure 3). For example, the 5th and 95 th percentile ranges of the risk distribution associated with modifiable factors were 2.9% to 5.0% and 15.5 to 25.0% for subjects who were in the lowest and highest deciles of nonmodifiable risk, respectively. Accordingly, estimates of the proportion of cases that could be prevented by the reduction of modifiable risks varied substantially across these strata, with a higher proportion of preventable cases in the strata defined by higher nonmodifiable risks (Table 2). In our model, we defined women at the lowest risk from modifiable risk factors as those who were in the lowest decile of BMI, did not use MHT, did not drink alcohol, and did not smoke. Overall, we estimated that up to 28.9% of all breast cancers could be prevented if all white women in the US population were at the lowest risk from these 4 modifiable risk factors. Nearly one-fifth of these total preventable cases arise from the subpopulation in the top decile of nonmodifiable risk. In contrast, only about 4% of the preventable cases arise from the population in the lowest decile of nonmodifiable risk.

Discussion

Utilizing a model including most known risk factors for breast cancer, we have shown that this information can be used to identify white women in the US population at substantially different levels of absolute risk for invasive breast cancer. We have also shown that the benefit (in terms of reductions in absolute risk) this population could achieve by changing modifiable risk factors is expected to be larger for those who are at higher than lower risk from nonmodifiable factors. This indicates that individual information on risk could be useful in making more informed decisions on breast cancer prevention.

Our results are generally consistent with the theoretical projections made regarding the degree of risk stratification achievable for various breast cancer risk models under a number of assumptions, including multiplicative effects of genetic and other risk factors.[10] Like other recent large studies,[18,19,21,29–32] we did not detect any evidence of multiplicative interactions between lifestyle and/or environmental risk factors and SNPs. Moreover, by application of a novel χ^2 goodness-of-fit test[22] designed to detect model misspecification at extremes of disease risk, our analysis provides additional evidence that a multiplicative model for geneenvironment interactions is adequate for describing the joint risk of breast cancer for women with different risk factor profiles. This was shown for the 24 SNPs that were genotyped in our sample. We could not validate the multiplicative assumption of the

model for the full set of 92 SNPs owing to the lack of genotyped data on 68 SNPs. However, our analyses of 24 SNPs and other very large, previously published studies[18,19,21,29–32] including more SNPs provide solid support for the multiplicative model. Multiplicative effects across many risk factors, even when individual effects are modest, can lead to pronounced stratification for absolute risk of breast cancer, as described in this report. The multiplicative model also implies that the absolute risk difference from modifiable risk factors varies by levels of nonmodifiable risk factors.[33]

The US Preventive Services Task Force currently recommends biennial screening mammography for women ages 50 to 74 years and consideration of individual factors, such as risk and potential benefit, for the decision to start screening mammography prior to age 50 years. Our analysis shows that use of a model based on most known risk factors can change the recommendation for screening for a substantial fraction of the population, compared with using only age-based criteria (Table 1). For example, a full model based on PRS and other risk factors can identify 16.1% of the population who can be recommended to start screening at age 40 years as their 10-year risk exceeds that of an average 50-year-old woman. However, the number of additional cases that would be detectable by screening would still be low, as a percentage of the women for whom risk needs to be assessed, and thus the population-level benefit of such practice, would depend on the implementation cost of risk assessment. The full model can also identify 32.0% of the population who at age 50 years have 10-year risk less than that of an average 40-year-old woman. These women benefit least from screening and may benefit from additional counseling about risk of false-positive results.

Results from these analyses could have implications for future cancer prevention efforts, particularly for risk communication and counseling at an individual level. For instance, women found to be at elevated risk owing to factors that cannot be changed may be more motivated to adopt a healthy lifestyle to lower their risk of breast cancer if they had a better understanding of the potential gains. In this regard, it is encouraging that even for women in the highest decile of risk owing to nonmodifiable factors, those who had low BMI, did not smoke or drink, and did not use MHT, had risks comparable to those for an average woman in the general population. Further research is needed to evaluate how knowledge of individual risk can influence behavior to modify risk.[34,35] Early studies[36–39] that have evaluated whether knowledge of genetic risk can improve health behavior have shown mixed results. As the number of susceptibility markers and their cumulative power to identify risk continue to increase for many common diseases, it will be increasingly important to develop and evaluate effective risk communication strategies that may motivate adoption of healthy behavior.

Consistent with a previous report from the United Kingdom,[4] our analysis indicates that only a modest proportion (29%) of breast cancer cases could be prevented by modifying most known risk factors. We also showed that a larger fraction of the total preventable cases would occur among women at higher levels of risk owing to genetic risk factors and other nonmodifiable risk factors. This could indicate that certain interventions for risk factor modification that may not be applicable to the whole population because of cost and other considerations could be targeted to high-risk strata to obtain a higher yield of cancers

prevented. As noted before, the cost-benefit ratio of such targeted intervention will depend on the cost of implementing risk assessment. However, a substantial proportion of cases preventable by modification of risk factors is still expected to arise outside the high-risk strata. Therefore, to have a major effect on reducing the disease burden, broader efforts for prevention need to continue at the population level. Furthermore, although these epidemiologic estimates of preventable cases could be a useful guide to understanding the potential effect of intervention and lifestyle change, ultimately evidence from randomized trials will be needed to understand the true effect of an intervention for the underlying population, as a whole or for subgroups.

Nonmodifiable risk factors were defined as those that cannot be modified (eg, genetics) or that are unlikely to be modified with the aim of reducing breast cancer risk. However, some of these factors do have modifiable components (height, age at menarche, and age at menopause are partially determined by diet and body size). A limitation of this report is that we could not evaluate several known risk factors for breast cancer since data were not available in the BPC3 data set. These include level of education, breastfeeding, physical activity, breast conditions (such as mammographic density and benign breast disease), and endogenous hormone biomarkers (such as estradiol, testosterone, and prolactin levels).[40] Further model improvements could also be achieved by refining the risk factors included in the model (eg, changes inBMI since age 18 years rather than current BMI). Our risk projections accounted for expected changes in MHT use over time based on the population distribution of length of use. However, our model assumed that all other risk factors remained constant over the time period of projected risk. Thus, the proportion of preventable cases including all known modifiable risk factors could be larger than reported here.

As information on all risk factors was not available in a single large study, we developed the model using a combination of imputation (for risk factors that were available in BP C3 but had missing data) and simulation (for PRS associated with 68 SNPs not genotyped in BPC3). Use of imputation within BPC3 allowed us to obtain more precise estimates of model parameters than those that could be obtained had we analyzed patients with only complete data. Nevertheless, when additional variation due to imputation was accounted for, substantial uncertainty in estimates of OR parameters was observed for several risk factors (eFigure 1 in the Supplement). In contrast, the use of simulation for 68 SNPs allowed us to incorporate information on very precise estimates of the OR parameters that are available from much larger case-control studies. In principle, risk estimates can be biased owing to the violation of the underlying assumption of multiplicative effect of SNPs and other risk factors, but for reasons noted earlier herein, this scenario is unlikely. As incidence density sampling was not followed in all studies, it is also possible that there could be some bias due to the use of ORs to estimate the hazard ratio parameters underlying the absolute risk model (eMethods in the Supplement). The effects of different types of biases owing to various modeling assumptions need to be examined in future validation studies.

Our analysis also has several strengths, including the development of a model for relative risks based on a large case-control sample drawn from prospective cohort studies, the incorporation of information on cancer rate and risk factor distributions from nationally representative databases, and the use of novel methodologic framework for assessment of

risk stratification. Future studies are needed to evaluate the value of incorporating additional information on factors into a model. Although our model assumptions are supported by analyses of very large sets of data, this model, as well as future extensions (eg, including more SNPs and other risk factors), need to be validated in independent prospective cohort studies. A more precise estimate ofrisk parameters associated with some of the epidemiologic risk factors could be used to reduce uncertainty in the estimates of risk that are produced by the model.

Conclusions

Our results illustrate the potential value of risk stratification to improve breast cancer prevention, particularly to aid decisions on risk factor modification at the individual level. The effect of such models for improving the cost-benefit ratio of population-based prevention programs will depend on the implementation cost of risk assessment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. National Cancer Institute. Incident Cases of Breast Cancer in 2014. 2014. http://www.cancer.gov/cancertopics/types/breast/. Accessed October 14, 2014

2. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer. 2013; 132(5):1133–1145. [PubMed: 22752881]

3. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015; 136(5):E359–E386. [PubMed: 25220842]

4. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br J Cancer. 2011; 105(suppl 2):S77–S81. [PubMed: 22158327]

5. Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the United States explained by well-established risk factors. J Natl Cancer Inst. 1995; 87(22):1681–1685. [PubMed: 7473816]

6. Michailidou K, Beesley J, Lindstrom S, et al. BOCS; kConFab Investigators; AOCS Group; NBCS; GENICA Network. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet. 2015; 47(4):373–380. [PubMed: 25751625]

7. Michailidou K, Hall P, Gonzalez-Neira A, et al. Breast and Ovarian Cancer Susceptibility Collaboration; Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON); kConFab Investigators; Australian Ovarian Cancer Study Group; GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013; 45(4):353–361. e1–e2. [PubMed: 23535729]

8. Nature Publishing Group. iCOGS Primer. 2013. http://www.nature.com/icogs/primers/. Accessed May 30, 2013

9. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015; 107(5):djv036. [PubMed: 25855707]

10. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. J Natl Cancer Inst. 2014; 106(11):dju305. [PubMed: 25392194]

11. National Cancer Institute. Division of Cancer Control and Population Sciences. OncoArray Network. http://epi.grants.cancer.gov/oncoarray/. Accessed January 30, 2015

12. Burton H, Chowdhury S, Dent T, Hall A, Pashayan N, Pharoah P. Public health implications from COGS and potential for risk stratification and screening. Nat Genet. 2013; 45(4):349–351. [PubMed: 23535723]

13. Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med. 2008; 358(26):2796–2803. [PubMed: 18579814]

14. Wacholder S, Hartge P, Prentice R, et al. Performance of common genetic variants in breast-cancer risk models. N Engl J Med. 2010; 362(11):986–993. [PubMed: 20237344]

15. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J Natl Cancer Inst. 2008; 100(14):1037–1041. [PubMed: 18612136]

16. Hüsing A, Canzian F, Beckmann L, et al. BPC3. Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status. J Med Genet. 2012; 49(9):601–608. [PubMed: 22972951]

17. Darabi H, Czene K, Zhao W, Liu J, Hall P, Humphreys K. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. Breast Cancer Res. 2012; 14(1):R25. [PubMed: 22314178]

18. 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. J Natl Cancer Inst. 2011; 103(16): 1252–1263. [PubMed: 21791674]

19. Joshi AD, Lindström S, Hüsing A, et al. Breast and Prostate Cancer Cohort Consortium (BPC3). Additive interactions between susceptibility single-nucleotide polymorphisms identified in genome-wide association studies and breast cancer risk factors in the Breast and Prostate Cancer Cohort Consortium. Am J Epidemiol. 2014; 180(10):1018–102. [PubMed: 25255808]

20. Hunter DJ, Riboli E, Haiman CA, et al. National Cancer Institute Breast and Prostate Cancer Cohort Consortium. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. Nat Rev Cancer. 2005; 5(12):977–985. [PubMed: 16341085]

21. 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. Hum Mol Genet. 2014; 23(19):5260–5270. [PubMed: 24895409]

22. Song M, Kraft P, Joshi AD, Barrdahl M, Chatterjee N. Testing calibration of risk models at extremes of disease risk. Biostatistics. 2015; 16(1):143–154. [PubMed: 25027274]

23. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). Underlying Cause of Death 1999–2011 on CDC WONDER Online Database, released 2014. Data are from the Multiple Cause of Death Files, 1999–2011 as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. http://wonder.cdc.gov/ucd-icd10.html. Accessed August 26, 2014

24. National Cancer Institute DSRPCSB. Underlying mortality data provided by NCHS Surveillance, Epidemiology, and End Results (SEER) Program. http://www.seer.cancer.gov) SEER*Stat Database: Mortality-All COD, Aggregated With County, Total U.S. (1969–2010) <Katrina/Rita Population Adjustment>–Linked to County Attributes—Total U.S., 1969–2011 Counties 2014

25. Chyba MM, Washington LR. Questionnaires from the National Health Interview Survey, 1985–89. Vital Health Stat 1. 1993; 31:1–412.

26. Parsons VL, Moriarity C, Jonas K, Moore TF, Davis KE, Tompkins L. Design and estimation for the national health interview survey, 2006–2015. Vital Health Stat 2. 2014; (165):1–53.

27. Centers for Disease Control and Prevention, US Department of Health and Human Services. National Health Interview Survey (NHIS) Public Use Data Release, NHIS Survey Description 2011. 2010. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2010/srvydesc.pdf. Accessed October 14, 2014

28. Centers for Disease Control and Prevention (CDC). National Health and Nutrition Examination Survey Questionnaire. 2010. http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm. Accessed April 7, 2016

29. Nickels S, Truong T, Hein R, et al. Genica Network; kConFab; AOCS Management Group. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. PLoS Genet. 2013; 9(3):e1003284. [PubMed: 23544014]

30. Milne RL, Herranz J, Michailidou K, et al. kConFab Investigators; Australian Ovarian Cancer Study Group; GENICA Network; TNBCC. 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. Hum Mol Genet. 2014; 23(7):1934–1946. [PubMed: 24242184]

31. Schoeps A, Rudolph A, Seibold P, et al. Identification of new genetic susceptibility loci for breast cancer through consideration of gene-environment interactions. Genet Epidemiol. 2014; 38(1):84–93. [PubMed: 24248812]

32. Rudolph A, Milne RL, Truong T, et al. kConFab Investigators; AOCS Group; GENICA-Network. Investigation of gene-environment interactions between 47 newly identified breast cancer susceptibility loci and environmental risk factors. Int J Cancer. 2015; 136(6):E685–E696. [PubMed: 25227710]

33. Petracci E, Decarli A, Schairer C, et al. Risk factor modification and projections of absolute breast cancer risk. J Natl Cancer Inst. 2011; 103(13):1037–1048. [PubMed: 21705679]

34. McBride CM, Koehly LM, Sanderson SC, Kaphingst KA. The behavioral response to personalized genetic information: will genetic risk profiles motivate individuals and families to choose more healthful behaviors? Annu Rev Public Health. 2010; 31:89–103. [PubMed: 20070198]

35. Khoury MJ, Clauser SB, Freedman AN, et al. Population sciences, translational research, and the opportunities and challenges for genomics to reduce the burden of cancer in the 21st century. Cancer Epidemiol Biomarkers Prev. 2011; 20(10):2105–2114. [PubMed: 21795499]

36. Bloss CS, Schork NJ, Topol EJ. Effect of direct-to-consumer genomewide profiling to assess disease risk. N Engl J Med. 2011; 364(6):524–534. [PubMed: 21226570]

37. Garcia-Closas M, Couch FJ, Lindstrom S, et al. Gene ENvironmental Interaction and breast CAncer (GENICA) Network; kConFab Investigators; Familial Breast Cancer Study (FBCS); Australian Breast Cancer Tissue Bank (ABCTB) Investigators. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet. 2013; 45(4):392–398. e1–e2. [PubMed: 23535733]

38. Voils CI, Coffman CJ, Grubber JM, et al. Does type 2 diabetes genetic testing and counseling reduce modifiable risk factors? a randomized controlled trial of veterans. J Gen Intern Med. 2015; 30(11):1591–1598. [PubMed: 25876740]

39. Christensen KD, Roberts JS, Zikmund-Fisher BJ, et al. REVEAL Study Group. Associations between self-referral and health behavior responses to genetic risk information [published online January 31, 2015]. Genome Med. 2015; 7(1):10.doi: 10.1186/s13073-014-0124-0 [PubMed: 25642295]

40. Tworoger SS, Zhang X, Eliassen AH, et al. Inclusion of endogenous hormone levels in risk prediction models of postmenopausal breast cancer. J Clin Oncol. 2014; 32(28):3111–311. [PubMed: 25135988]

Key Points

Questions

What is the utility of low penetrant common single nucleotide polymorphisms (SNPs) for guiding public health strategies for breast cancer prevention?

Findings

A risk prediction model including 92 susceptibility SNPs and various epidemiologic risk factors can provide important stratification for absolute risk for white women in the United States. The model predicts that effect of healthy lifestyle choices for risk reduction is expected to be larger for women who are at higher risk owing to genetic susceptibility and other nonmodifiable risk factors.

Meaning

The assessment of common SNPs may be useful for screening recommendations and individualized risk communication.
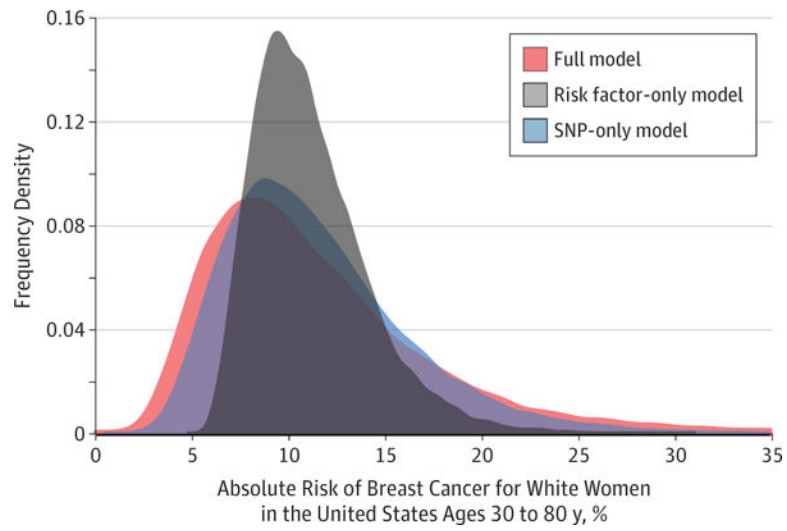
Figure 1. Projected Distribution of Absolute Lifetime Risk of Breast Cancer for White Women in the United States Ages 30 to 80 Years

SNP indicates single nucleotide polymorphisms.

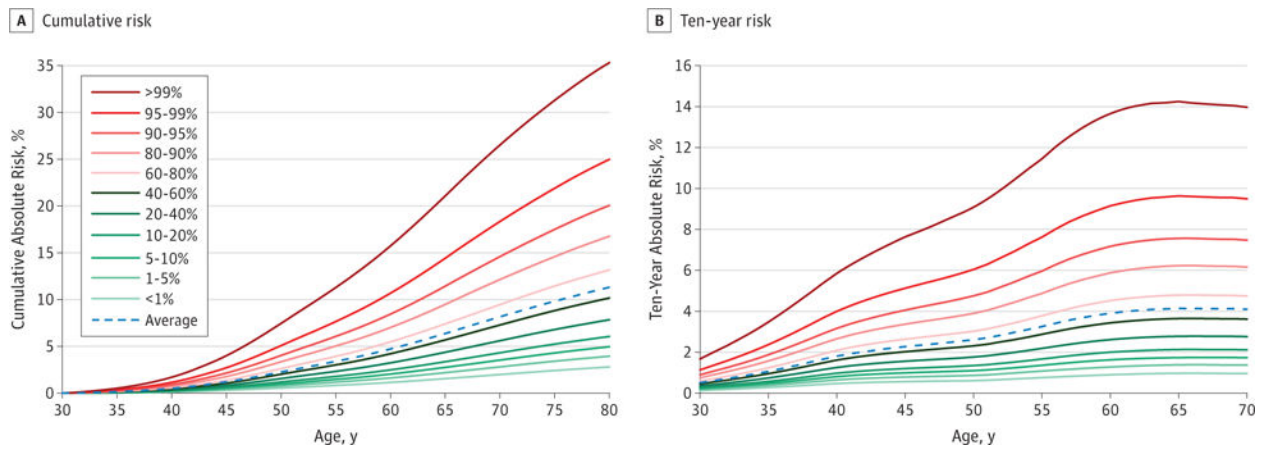**Figure 2. Cumulative and 10-Year Breast Cancer Risk for White Women in the United States Stratified by Risk Percentiles**

Cumulative risk is evaluated as absolute risk between age 30 years and a specific age shown on the x-axis. The 10-year risk is evaluated as absolute risk over the next 10 years for a woman who has attained a specific age (shown on the x-axis) without developing breast cancer.
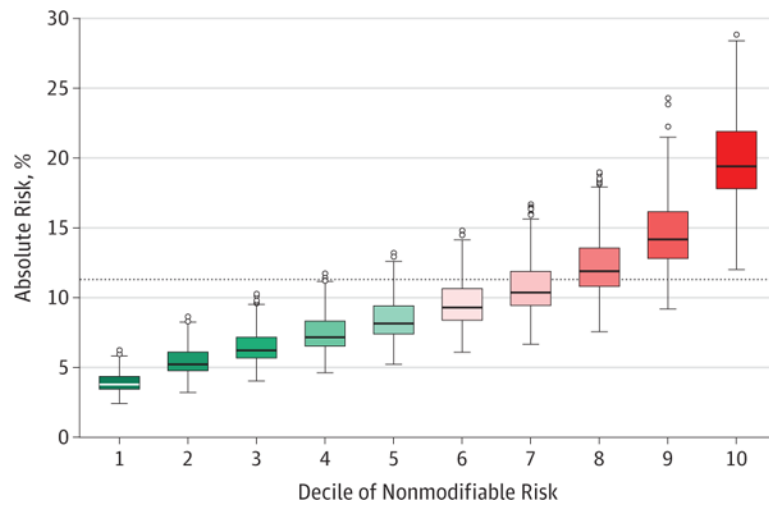
Figure 3. Distribution of Absolute Lifetime Risk Associated With Modifiable Risk Factors Stratified by Deciles of Nonmodifiable Risk for White Women in the United States
The horizontal line in the middle of each box indicates the median, while the top and bottom borders of the box mark the 75th and 25th percentiles, respectively. The whiskers above and below the box are the minimum and maximum excluding outliers; outliers were defined as individuals who had risk beyond above or below a standard deviation of 3 of means in the log-scale. Lifetime risk refers to cumulative risk between age 30 to 80 years. The dashed line indicates average lifetime risk for the population.

Table 1

Total Number of At-Risk Subjects and Incident Cases Expected at Different Risk Levels for Every 100 000 Women With Assessed Risk

Model						
	PRS-92 Only		**Questionnaire-Based Risk Factors Only**		**PRS-92 and Risk Factors**	
Risk Level	**Total Subjects, No.**	**Cases, No.**	**Total Subjects, No.**	**Cases, No.**	**Total Subjects, No.**	**Cases, No.**
Moderate risk: RR = 2–3[a]	2691	688	306	74	4116	1076
High risk: RR>3[a]	109	40	0	0	649	181
10-y risk at 40 is > average 10-y risk at 50[b]	9113	295	6531	194	16 134	564
10-y risk at 50 is < average 10-y risk at 40[c]	27 018	380	11 231	184	32 037	425

Abbreviations: PRS, polygenic risk score; PRS-92, all 92 known breast cancer SNPs; RR, relative risk; SNP, single nucleotide polymorphisms.

[a]The reference is 11.3%, the average risk in women ages 30 to 80 years.

[b]The average 10-y risk at age 50 years is 2.6%.

[c]The average 10-y risk at age 40 years is 1.8%.

Table 2

Estimates of Proportion of Breast Cancer Cases Preventable by Reduction of Modifiable Risk in Different Strata of the Population Defined by Nonmodifiable Risk Factors[a]

	Proportion of Breast Cancer, %									
	Alcohol		MHT		BMI[b]		Smoking		All 4 Modifiable Risk Factors Simultaneously[c]	
Nonmodifiable Risk Groups	P	T	P	T	P	T	P	T	P	T
Decile										
1	4.00	0.36	4.60	0.31	4.80	0.57	4.10	0.12	4.40	1.28
2	5.50	0.49	5.80	0.38	6.30	0.76	5.70	0.17	5.90	1.70
3	6.60	0.59	7.00	0.47	7.20	0.87	6.80	0.21	7.00	2.01
4	7.70	0.69	8.30	0.55	8.10	0.98	7.90	0.24	8.00	2.31
5	8.60	0.77	8.80	0.58	9.10	1.09	8.70	0.27	8.80	2.55
6	9.90	0.89	9.50	0.63	10.10	1.22	9.60	0.30	9.80	2.84
7	11.10	1.00	11.10	0.74	10.90	1.32	10.80	0.33	11.00	3.18
8	12.40	1.11	12.00	0.80	12.10	1.46	12.50	0.38	12.20	3.53
9	14.70	1.32	14.30	0.95	13.80	1.66	15.20	0.47	14.30	4.14
10	19.7	1.78	18.50	1.23	17.50	2.11	18.80	0.58	18.50	5.35
PAR[d]	–	9.01	–	6.64	–	12.05	–	3.08	–	28.90

Abbreviations: BMI, body mass index; MHT, menopausal hormone therapy; P, total number of preventable breast cancers; PAR, population-attributable risk; T, total number of breast cancers.

[a]The proportions for each stratum are shown relative to the total number of breast cancers (%T) and total number of preventable breast cancers (%P) that are expected to arise in the whole population.

[b]BMI is calculated as weight in kilograms divided by height in meters squared.

[c]The modifiable riskfactors are body mass index, MHT use, alcohol use, and smoking.

[d]Estimate of population-attributable risk due to modifiable factors (individually and simultaneously). PAR is given by column sum of T and %P = (%T/PAR) ×100.