Published in final edited form as:
N Engl J Med. 2010 March 18; 362(11): 986–993. doi:10.1056/NEJMoa0907727.

Performance of Common Genetic Variants in Breast-Cancer Risk Models

Sholom Wacholder, Ph.D., **Patricia Hartge, Sc.D.**, **Ross Prentice, Ph.D.**, **Montserrat Garcia-Closas, M.D., Ph.D.**, **Heather Spencer Feigelson, Ph.D.**, **W. Ryan Diver, M.S.P.H.**, **Michael J. Thun, M.D.**, **David G. Cox, Ph.D.**, **Susan E. Hankinson, Ph.D.**, **Peter Kraft, Ph.D.**, **Bernard Rosner, Ph.D.**, **Christine D. Berg, M.D.**, **Louise A. Brinton, Ph.D.**, **Jolanta Lissowska, Ph.D.**, **Mark E. Sherman, M.D.**, **Rowan Chlebowski, M.D.**, **Charles Kooperberg, Ph.D.**, **Rebecca D. Jackson, M.D.**, **Dennis W. Buckman, Ph.D.**, **Peter Hui, B.S.**, **Ruth Pfeiffer, Ph.D.**, **Kevin B. Jacobs, B.S.**, **Gilles D. Thomas, M.D.**, **Robert N. Hoover, M.D., Sc.D.**, **Mitchell H. Gail, M.D., Ph.D.**, **Stephen J. Chanock, M.D.**, and **David J. Hunter, M.B., B.S., Sc.D.**
Division of Cancer Epidemiology and Genetics (S.W., P.H., M.G.-C., L.A.B., M.E.S., R. Pfeiffer, K.B.J., G.D.T., R.N.H., M.H.G., S.J.C.) and the Early Detection Research Group, Division of Cancer Prevention (C.D.B.), National Cancer Institute, Bethesda, MD; the Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle (R. Prentice, C.K.); the Department of Epidemiology, American Cancer Society, Atlanta (H.S.F., W.R.D., M.J.T.); the Institute for Health Research, Kaiser Permanente, Denver (H.S.F.); INSERM, Lyon, France (D.G.C.); the Departments of Epidemiology (D.G.C., S.E.H., P.K., D.J.H.) and Biostatistics (B.R.), Harvard School of Public Health; and Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School (S.E.H., B.R., D.J.H.) — all in Boston; the Department of Cancer Epidemiology and Prevention, Maria Sklodowska- Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland (J.L.); the Division of Medical Oncology and Hematology, Los Angeles Biomedical Research Institute at Harbor– UCLA Medical Center, Torrance, CA (R.C.); Ohio State University, Columbus (R.D.J.); and Information Management Services, Silver Spring, MD (D.W.B., P.H.)

Abstract

BACKGROUND—Genomewide association studies have identified multiple genetic variants associated with breast cancer. The extent to which these variants add to existing risk-assessment models is unknown.

METHODS—We used information on traditional risk factors and 10 common genetic variants associated with breast cancer in 5590 case subjects and 5998 control subjects, 50 to 79 years of age, from four U.S. cohort studies and one case–control study from Poland to fit models of the absolute risk of breast cancer. With the use of receiveroperating- characteristic curve analysis, we calculated the area under the curve (AUC) as a measure of discrimination. By definition, random classification of case and control subjects provides an AUC of 50%; perfect classification provides an AUC of 100%. We calculated the fraction of case subjects in quintiles of estimated absolute risk after the addition of genetic variants to the traditional risk model.

RESULTS—The AUC for a risk model with age, study and entry year, and four traditional risk factors was 58.0%; with the addition of 10 genetic variants, the AUC was 61.8%. About half the

Address reprint requests to Dr. Wacholder at the Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd., EPS 5050, MSC-7244, Bethesda, MD 20892, or at wacholds@mail.nih.gov.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

case subjects (47.2%) were in the same quintile of risk as in a model without genetic variants; 32.5% were in a higher quintile, and 20.4% were in a lower quintile.

CONCLUSIONS—The inclusion of newly discovered genetic factors modestly improved the performance of risk models for breast cancer. The level of predicted breast-cancer risk among most women changed little after the addition of currently available genetic information.

Personalized medicine, the assignment of preventive measures or treatment interventions on the basis of individual characteristics, can result in better outcomes than the use of the same strategy for everyone. Recent changes in the U.S. Preventive Services Task Force guidelines[1] for mammographic screening raise the question of whether recommendations about age at the onset of screening and the frequency of screening can be calibrated to an individual woman's risk of breast cancer. Clinicians already use guidelines in making decisions about assessments to identify carriers of rare *BRCA1* and *BRCA2* mutations, which confer very high risks of breast cancer and ovarian cancer.[2] Single-nucleotide polymorphisms (SNPs) that have been shown to be associated with breast cancer in genomewide association studies[3] are common but confer only small increases in risk. Whether these variants collectively enhance the identification of women at increased or reduced risk is unknown.

The Breast Cancer Risk Assessment Tool,[4,5] commonly referred to as the Gail model, summarizes information about a woman's reproductive history, breast cancer in close relatives, and previous breast biopsies to estimate the probability of the development of breast cancer in subsequent years; the estimated breast-cancer risk based on the Gail model has been used for counseling, informing decisions about the use of tamoxifen,[6] and determining sample size in randomized prevention trials.[7] In this study, we empirically evaluated the contribution of a set of 10 newly established common genetic variants as an alternative to and as a supplement to the components of the Gail model in 5590 case subjects and 5998 control subjects, 50 to 79 years of age, included in the National Cancer Institute's Cancer Genetic Markers of Susceptibility multistage genomewide association study of breast cancer.[8]

METHODS

STUDIES

Investigators from the Women's Health Initiative Observational Study,9 the American Cancer Society Cancer Prevention Study II Nutrition Cohort,[10] the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (ClinicalTrials.gov number, NCT00002540),[11] and the Nurses' Health Study12 collected information and specimens at baseline and clinical observations of the subjects for up to 15 years of follow-up. We also used data from the Polish Breast Cancer Study,13 a population-based case–control study in Warsaw and Lodz, Poland.

SUBJECTS

All case subjects were women who had received a diagnosis of invasive breast cancer. Additional descriptions of the studies, including methods for case ascertainment and selection of control subjects, are included in Appendixes 1, 2, and 3 in the Supplementary Appendix, available with the full text of this article at NEJM.org.

COMPONENTS OF THE GAIL MODEL

We included components of the Gail model that were collected prospectively in the four cohort studies and retrospectively in the case–control study. These components were the number of first-degree relatives with a diagnosis of breast cancer, age at menarche, age at

first live birth, and number of previous breast biopsies. Mammographic density and diagnosis of atypical hyperplasia, which have improved the performance of the Gail model in other studies,[4,14] were not available for most studies and were not used in this analysis.

GENETIC VARIANTS AND GENOTYPING

We analyzed genotypes in case subjects and control subjects for 10 SNPs with established associations with breast cancer that achieved genomewide significance in three published genomewide association studies.[8,15] Some of the data used in our analysis were reported in earlier articles on the discovery of susceptibility alleles.[8,16-18] Genotypes used in this analysis were described in a recent article about the Cancer Genetic Markers of Susceptibility study.[8] Complete genotype data from a total of 5590 case subjects and 5998 control subjects were available for analysis after the exclusions described in Table 1 in the Supplementary Appendix.

MODELS

We fit a family of five logistic-regression models, including combinations of demographic factors (age, entry year, and cohort), the components of the Gail model, and genotypes for the 10 SNPs. The demographic model included demographic factors only; adjustment for age is described in Appendix 4 in the Supplementary Appendix. The nongenetic model added components of the Gail model to incorporate standard risk factors and reflect current clinical practice. Two models included demographic and genetic factors only: the genetic variant-count model included the number of risk-conferring variant alleles (one allele for heterozygotes and two alleles for homozygotes) in categories of fewer than 6, 7 to 8, 9 to 10, 11 to 12, and 13 or more; the genetic individual-variant model included separate effects of the 10 individual SNPs. The inclusive model combined demographic factors, components of the Gail model, and the 10 SNPs into a single model. Tables 1, 2, and 3 show the distribution of case subjects and control subjects according to cohort, age, and factors in the models. Except for age and year of study entry, all variables were fit as categorical terms.

INTERACTIONS

We investigated whether interactions between components of the Gail model and genetic variants would improve the performance of the relevant models. To avoid complex issues involving model selection that are beyond the scope of this article, we evaluated nonmultiplicative effects using the Tukey 1-degree-of-freedom test applied to the interaction term between the risk scores based on the Gail model and those based on the genetic variants[19,20] and a 16-degree-of-freedom test based on quintiles of the two scores. Further details are included in Appendix 5 in the Supplementary Appendix.

ESTIMATES OF ABSOLUTE RISK

We combined the risks from the logistic models, crude rates of breast cancer for each underlying cohort, and registry data to estimate the absolute risk of breast cancer per year, as described in Appendix 6 in the Supplementary Appendix. We used rates of breast cancer from U.S.[21] and Polish[22] registries, as appropriate.

MODEL DISCRIMINATION

Using receiver-operating-characteristic (ROC) curve analysis, we calculated the area under the curve (AUC), also known as a concordance (C) statistic, to assess discrimination. Each point on the ROC curve shows the effect of a rule for turning a risk estimate into a prediction of the development of breast cancer in a woman. The y axis of the ROC curve is the true positive rate or sensitivity (i.e., the proportion of women with breast cancer who were correctly predicted to have the disease). The x axis shows the false positive rate (the

complement of specificity) (i.e., the proportion of women without breast cancer who were incorrectly predicted to have breast cancer). The area under the ROC curve, the AUC, measures how well the model discriminates between case subjects and control subjects. An ROC curve that corresponds to a random classification of case subjects and control subjects is a straight line with an AUC of 50%. An ROC curve that corresponds to perfect classification has an AUC of 100%. Figure 1 shows a modified ROC graph, in which the y axis is the true positive rate and the x axis is the false positive rate in the population instead of the rate among control subjects only.[15]

COMPARISON OF PERFORMANCE OF RISK MODELS

To compare the performance of the models, we calculated the cross-classification of the assignment of case subjects into predicted risk bands of women with similar risk. The cutoff points were determined according to quintiles of estimated risk among case subjects and control subjects, weighted according to the inverse sampling fraction so as to apply the model to the underlying cohort (Appendix 7 in the Supplementary Appendix). We also used integrated discrimination improvement[23] as a measure to compare models without setting cutoff points. This measure compares two models for the change in the difference in the average estimated annual risk of breast cancer between case subjects and control subjects[20]; thus, the integrated discrimination improvement will be large for a model that assigns a greater estimated risk to case subjects and a lesser estimated risk to control subjects. Details about the assessment of precision, testing of models, and extent of overfitting are included in Appendixes 8 and 9 in the Supplementary Appendix.

RESULTS

The associations between each predictor and breast-cancer risk were consistent with published data, and they were similar in the different models (Table 2 in the Supplementary Appendix). Notably, the effect of a family history of breast cancer did not materially change after adjustment for SNPs. The risk of breast cancer among women who carried 13 or more of a maximum of 20 riskconferring variant alleles (which occur in about 4% of the population) was nearly three times the risk among women who carry 6 or fewer variants (12%).

Table 3 in the Supplementary Appendix shows a comparison of how the models discriminated between women with breast cancer and those without the disease. The AUC for the demographic model was 53.4%. The addition of the SNP from the fibroblast growth factor receptor 2 gene (*FGFR2*) increased the AUC to 55.7%. The addition of four more SNPs (from Tox high-mobility group box family member 3 [*TOX3*], mitogenactivated protein kinase kinase kinase 1 [*MAP3K1*], chromosome 1p, and a RAD51-like 1 [*RAD51L1*]) increased the AUC further, to 57.9%. The AUC for the 7 SNPs described by Pharoah et al.[15] was 58.6%, and the addition of the 3 others to incorporate all 10 SNPs listed in Table 3 in the Supplementary Appendix increased the AUC only to 59.7%. The discrimination in the simpler genetic variant-count model was somewhat less (58.9%).

Among the individual nongenetic factors, a history of breast biopsy yielded the greatest AUC (56.2%). The nongenetic model, with the use of all four components of the Gail model, showed an AUC of 58.0%. The inclusive model, with components of the Gail model and 10 SNPs, yielded an AUC of 61.8%, an increase of about 3.8 percentage points over the nongenetic model and 2.1 percentage points over the genetic individualvariant model. In short, the AUC for the genetic models was only slightly higher than the AUC for the nongenetic model, and the addition of the components from the Gail model increased the AUC by 2 more percentage points; thus, discrimination by means of the inclusive model is better than discrimination by means of either the genetic or nongenetic models. In parallel,

integrated discrimination improvement, or the difference in the estimated annual risk of breast cancer between case subjects and control subjects, ranged from 0.015 to 0.022% in the inclusive model (Table 3 in the Supplementary Appendix). Our empirical analysis indicated a modest increment in discrimination by means of the inclusive model in relative and absolute terms, despite very low P values.

We also empirically examined the performance of high and low absolute risk (i.e., the performance of the models for women who might be singled out for more aggressive or less aggressive intervention than women at average risk). The genetic models yielded better discrimination than the nongenetic models among women at the lowest estimated risk (Fig. 1B), but not among women at the highest estimated risk (Fig. 1C). For women with a risk estimated from the inclusive model that was above the median in control subjects, the AUC was 58.8%, but it was only 56.6% for women with an estimated risk that was above the 80th percentile.

Table 4 in the Supplementary Appendix shows how the addition of genetic data affected classification of the case subjects with breast cancer. The percentages of all case subjects according to quintiles of absolute risk are shown for the inclusive and nongenetic models. The estimated annual risk was above 0.575% for substantially more case subjects with the inclusive model than with the nongenetic model (27.7% vs. 18.9%). Slightly fewer than half the case subjects (47.2%) were in the same category in both models. A total of 32.5% of the case subjects were in a higher-risk category with the inclusive model than with the nongenetic model, and 20.4% were in a lower category. The corresponding percentages for the control subjects were 26% and 28%. We found no evidence that a complex model involving interactions among Gail model components and SNPs performed materially better than the models described here ($P>0.10$).

DISCUSSION

In our study involving 5590 case subjects with breast cancer and 5998 control subjects, the addition of information on 10 genetic variants to a standard clinical breast-cancer risk model predicted the risk of breast cancer only slightly better than the clinical model alone. In the inclusive model, the ROC curve was above 60% of the possible AUC. That is, about 60% of the time, a randomly selected patient with breast cancer had a higher estimated risk than the risk for a randomly selected woman in whom breast cancer did not develop during the follow-up period. By contrast, a single dichotomous risk factor detected in 60% of case subjects and 40% of control subjects (odds ratio, 2.25; AUC, 60%) would discriminate about as well as our model by the AUC criterion. We saw similarly modest improvement using measures based on the change in estimated risk. Although our data suggest that the SNP-only model predicted risk slightly better than the Gail model, the nongenetic clinical variables are available at essentially no cost, whereas the costs of obtaining genetic information are likely to be substantial.

We have evaluated risk-prediction models as we would any other clinical test. Our results are presented in terms of absolute risks, which are easily translated into positive and negative predictive values in considering who will benefit from a clinical test, including a risk model.[24-26] Our findings were based on a large number of case subjects and control subjects, drawn from four prospective cohorts and one population-based case–control study. Our inclusive model incorporated information on components of the Gail model and genotypes of newly established SNPs; this information is not often included together in the same study. Our analysis also had technical advantages over other risk-modeling efforts: we fitted the effects of age and cohort so as not to give the models credit for fit based on

demographic factors; we were able to assess the effect of adding SNPs on estimated risk and on changes in risk category for case subjects and for a hypothetical population.

There were some unavoidable weaknesses of this study and its data sources, particularly in evaluating risk for clinical use. For example, we pooled data from four U.S. cohort studies and a case–control study from Poland; these studies had different designs and enrollment characteristics. We included only women of European ancestry in the empirical analysis, and we did not consider subtypes of breast cancer. As a group, the study participants are not representative of any specific population. Screening practice may vary within and among studies. Because of self-selection for participation in the studies, we expect that the average risks in the U.S. studies will be different from each other and from the U.S. average. Similarly, in a clinical context, we cannot expect the risk for an individual woman to be the same as the average risk for a population. Our estimates of the performance of the risk models may be slightly higher than can be expected in typical clinical settings because we report results with minor overfitting. Given the complexity of the model with 4 Gail model components and 10 SNPs, we chose to fit each factor without constraints but did not attempt to evaluate hundreds of interactions among the factors or with age and cohort. Our simple interaction models gave no suggestion of improved performance from interactions between genetic and Gail model components; more sophisticated modeling may improve the performance of these factors in predicting breast-cancer risk. Finally, comparisons of models are slightly unfavorable to the Gail model because we could not include the history of atypical hyperplasia and information on mammographic density. Although they are not routinely incorporated into the Gail model, these factors may improve the performance of future versions of the model.[14]

Our analysis indicates that the genetic variants we studied provide modest improvements in discrimination and prediction models, whether measured as the AUC, as a discrimination index, or as a change in position in broad bands of risk, such as might be used in clinical settings. We see little evidence of benefit from including genetic variants at the extremes of high and low risk, categories in which further stratification might be most valuable. This empirical demonstration of the potential benefit of adding SNP data to breast-cancer risk models based on individual data is generally consistent with theoretical predictions[15,27] that use published estimates of effect.

Because statistics such as the AUC and integrated discrimination improvement do not provide a readily intuitive sense of the clinical usefulness of these models, we focused much of our discussion on the degree of incremental improvement associated with adding the genetic variants. As in diabetes[28] and cardiovascular disease,[29] the addition of the common SNPs added little to the predictive value of the clinical models. On the basis of theoretical models, Gail[30] has shown that increases in the AUC similar to those observed here are not sufficiently large to improve meaningfully the identification of women who might benefit from tamoxifen prophylaxis or assignment of screening mammography.

Although the Gail model and SNP-inclusive models may help to identify groups of women who have an increased risk of breast cancer for trials of interventions, none of the models in our set of data accurately predicted the development of breast cancer. Our results indicate that the recent identification of common genetic variants does not herald the arrival of personalized prevention of breast cancer in most women. Even with the addition of these common variants, breast-cancer risk models are not yet able to identify women at reduced or elevated risk in a clinically useful way.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported in part by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics of the National Cancer Institute and by grants from the National Institutes of Health (N01WH22110, P01 CA87969, P01 CA53996, R01 CA49449, U01CA098233, and U01CA098710).

We thank Prisca Fall-Keita and David Check for assistance in preparation of an earlier version of the manuscript and the late Dr. Eugenia Calle for her contributions in developing the American Cancer Society cohort.

References

1. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009;151:716–26. [PubMed: 19920272]
2. ACOG practice bulletin no. 103: hereditary breast and ovarian cancer syndrome. Obstet Gynecol 2009;113:957–66. [PubMed: 19305347]
3. Manolio TA, Brooks LD, Collins FSA. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008;118:1590–605. [PubMed: 18451988]
4. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879–86. [PubMed: 2593165]
5. Gail MH, Benichou J. Validation studies on a model for breast cancer risk. J Natl Cancer Inst 1994;86:573–5. Erratum, J Natl Cancer Inst 1994;86:803. [PubMed: 8179704]
6. Gail MH, Costantino JP, Bryant J, et al. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. J Natl Cancer Inst 1999;91:1829–46. Erratum, J Natl Cancer Inst 2000;92:275. [PubMed: 10547390]
7. Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst 1999;91:1541–8. [PubMed: 10491430]
8. Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009;41:579–84. [PubMed: 19330030]
9. Langer RD, White E, Lewis CE, Kotchen JM, Hendrix SL, Trevisan M. The Women's Health Initiative Observational Study: baseline characteristics of participants and reliability of baseline measures. Ann Epidemiol 2003;13(Suppl):S107–S121. [PubMed: 14575943]
10. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer 2002;94:2490–501. [PubMed: 12015775]
11. Hayes RB, Reding D, Kopp W, et al. Etiologic and early marker studies in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials 2000;21(Suppl): 349S–355S. [PubMed: 11189687]
12. Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Hennekens CH, Speizer FE. Moderate alcohol consumption and the risk of breast cancer. N Engl J Med 1987;316:1174–80. [PubMed: 3574368]
13. García-Closas M, Egan KM, Newcomb PA, et al. Polymorphisms in DNA doublestrand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. Hum Genet 2006;119:376–88. [PubMed: 16485136]
14. Chen JB, Pee D, Ayyagari R, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst 2006;98:1215–26. [PubMed: 16954474]
15. Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med 2008;358:2796–803. [PubMed: 18579814]
16. Ahmed S, Thomas G, Ghoussaini M, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 2009;41:585–90. [PubMed: 19330027]

17. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93. [PubMed: 17529967]
18. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4. [PubMed: 17529973]
19. Tukey JW. One degree of freedom for non-additivity. Biometrics 1949;5:232–42.
20. Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. Am J Hum Genet 2006;79:1002–16. [PubMed: 17186459]
21. Surveillance Epidemiology and End Results (SEER) program. Bethesda, MD: National Cancer Institute; 2007 [February 11, 2010]. SEER 17 incidence and mortality, 2000-2003, with Kaposi sarcoma and mesothelioma. at http://www.seer.cancer.gov/seerstat
22. Number of new cancer cases by site and age groups. The Maria Skłodowska-Curie Memorial Cancer Center, Department of Epidemiology and Cancer Prevention. [February 11, 2010]. at http://epid.coi.waw.pl/krn/english/liczba_zach_rozp/default.asp
23. Pencina MJ, D'Agostino RB, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72. 207–12. [PubMed: 17569110]
24. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. Lancet 2007;370:890–907. [PubMed: 17826171]
25. Kraft P, Wacholder S, Cornelis MC, et al. Beyond odds ratios — communicating disease risk based on genetic profiles. Nat Rev Genet 2009;10:264–9. [PubMed: 19238176]
26. Katki HA, Wacholder S, Solomon D, Castle PE, Schiffman M. Risk estimation for the next generation of prevention programmes for cervical cancer. Lancet Oncol 2009;10:1022–3. [PubMed: 19767237]
27. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J Natl Cancer Inst 2008;100:1037–41. [PubMed: 18612136]
28. Lyssenko V, Jonsson A, Almgren P, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med 2008;359:2220–32. [PubMed: 19020324]
29. Kathiresan S, Melander O, Anevski D, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med 2008;358:1240–9. [PubMed: 18354102]
30. Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. J Natl Cancer Inst 2009;101:959–63. [PubMed: 19535781]

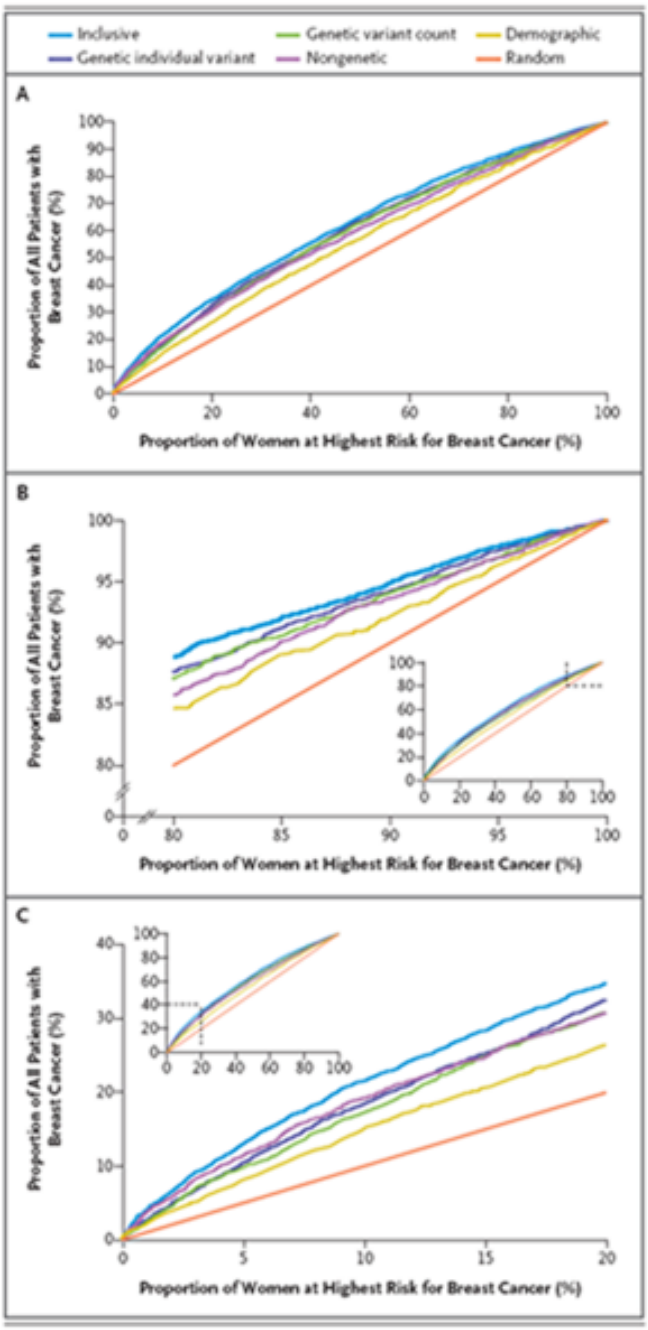

Figure 1. Five Models of Breast-Cancer Risk

Data are from a hypothetical population comprising the five study populations. The straight line indicates random classification. The proportions on the curve run from highest to lowest. For example, in Panel A, the inclusive model indicates that 34% of women who had breast cancer were among the 20% of women at highest risk. Panel B shows results limited to women in the lowest 20% of estimated risk, according to the inclusive model. The inset shows where this lowest 20% would be located in Panel A. Panel C shows results limited to women in the highest 20% of estimated risk according to the inclusive model. The inset shows where this highest 20% would be located in Panel A.

Table 1

Baseline Data for the Subjects.*

Variable	Control Subjects (N = 5998)	Case Subjects (N = 5590)	All Subjects (N = 11,588)
		number (percent)	
Age group			
50–59 yr	1155 (19.3)	1181 (21.1)	2336 (20.2)
60–69 yr	2655 (44.3)	2544 (45.5)	5199 (44.9)
70–79 yr	2188 (36.5)	1865 (33.4)	4053 (35.0)
Study			
Nurses' Health Study	948 (15.8)	1005 (18.0)	1953 (16.9)
Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial	898 (15.0)	730 (13.1)	1628 (14.0)
Women's Health Initiative Observational Study	2216 (36.9)	2205 (39.4)	4421 (38.2)
American Cancer Society Cancer Prevention Study II Nutrition Cohort	506 (8.4)	388 (6.9)	894 (7.7)
Polish Breast Cancer Study	1430 (23.8)	1262 (22.6)	2692 (23.2)

* Percentages may not total 100 because of rounding.

Table 2

Risk Factors Associated with Breast Cancer.*

Variable	Control Subjects (N = 5998)	Case Subjects (N = 5590)	All Subjects (N = 11,588)
		number (percent)	
Age at menarche			
≥14 yr	1842 (30.7)	1626 (29.1)	3468 (29.9)
12–13 yr	3102 (51.7)	2860 (51.2)	5962 (51.4)
7–11 yr	1054 (17.6)	1104 (19.7)	2158 (18.6)
No. of biopsies			
0	4897 (81.6)	4135 (74.0)	9032 (77.9)
1	865 (14.4)	1179 (21.1)	2044 (17.6)
≥2	236 (3.9)	276 (4.9)	512 (4.4)
Age at first live birth			
<20 yr	1056 (17.6)	909 (16.3)	1965 (17.0)
20–24 yr	2675 (44.6)	2214 (39.6)	4889 (42.2)
25–29 yr, or no births	1803 (30.1)	1918 (34.3)	3721 (32.1)
≥30 yr	464 (7.7)	549 (9.8)	1013 (8.7)
No. of first-degree relatives with breast cancer			
0	5188 (86.5)	4567 (81.7)	9755 (84.2)
1	735 (12.3)	918 (16.4)	1653 (14.3)
≥2	75 (1.3)	105 (1.9)	180 (1.6)
No. of risk-conferring variant alleles			
0–6	731 (12.2)	432 (7.7)	1163 (10.0)
7 or 8	1868 (31.1)	1355 (24.2)	3223 (27.8)
9 or 10	2164 (36.1)	2064 (36.9)	4228 (36.5)
11 or 12	1001 (16.7)	1341 (24.0)	2342 (20.2)
≥13	234 (3.9)	398 (7.1)	632 (5.5)

* Percentages may not total 100 because of rounding.

Table 3

Common Genetic Variants Associated with Breast Cancer.*

SNP	Chromosome	Gene	High-Risk Allele	Low-Risk Allele	Frequency of High-Risk Allele
					percent
RS1045485	2q	*CASP8*	G	C	86.7
RS13281615	8q	Unknown	G	A	41.3
RS13387042	2q	Unknown	A	G	52.1
RS2981582	10q	*FGFR2*	T	C	38.1
RS3803662	10q	*TOX3*	T	C	27.1
RS3817198	11p	*LSP1*	C	T	32.5
RS889312	16q	*MAP3K1*	C	A	27.8
RS7716600	5p	Unknown	A	C	22.0
RS11249433	1p	Unknown	C	T	39.3
RS999737	14q	*RAD51L1*	C	T	76.3

**CASP8* denotes caspase 8, apoptosis-related cysteine peptidase, *FGFR2* fibroblast growth factor receptor 2, *LSP1* lymphocyte-specific protein 1, *MAP3K1* mitogen-activated protein kinase kinase kinase 1, *RAD51L1*, RAD51-like 1, SNP single-nucleotide polymorphism, and *TOX3* Tox high-mobility group box family member 3.

N Engl J Med. Author manuscript; available in PMC 2010 September 1.