



Published in final edited form as:

*Cancer Epidemiol Biomarkers Prev.* 2021 April ; 30(4): 600–607. doi:10.1158/1055-9965.EPI-20-0900.

## Simplified breast risk tool integrating questionnaire risk factors, mammographic density, and polygenic risk score: development and validation

**Bernard Rosner<sup>1,2</sup>, Rulla M. Tamimi<sup>1,3,4</sup>, Peter Kraft<sup>4,8</sup>, Chi Gao<sup>8</sup>, Yi Mu<sup>1</sup>, Christopher Scott<sup>5</sup>, Stacey J. Winham<sup>5</sup>, Celine M. Vachon<sup>6</sup>, Graham A. Colditz<sup>7</sup>**

<sup>1</sup>Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

<sup>2</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

<sup>3</sup>Division of Epidemiology, Population Health Sciences Department, Weill Cornell Medicine, New York

<sup>4</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

<sup>5</sup>Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota

<sup>6</sup>Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota

<sup>7</sup>Alvin J. Siteman Cancer Center and Department of Surgery, Division of Public Health Sciences, School of Medicine, Washington University in St. Louis, St. Louis, Missouri

<sup>8</sup>Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

### Abstract

**Background:** Clinical use of breast cancer risk prediction requires simplified models. We evaluate a simplified version of the validated Rosner-Colditz model and add percent mammographic density (MD) and polygenic risk score (PRS), to assess performance from ages 45–74. We validate using the Mayo Mammography Health Study (MMHS).

**Methods:** We derived the model in the Nurses' Health Study (NHS) based on: MD, 77SNP PRS and a questionnaire score (QS) (lifestyle and reproductive factors). 2799 invasive breast cancer cases were diagnosed from 1990–2000. MD (using Cumulus software) and PRS were assessed in a nested case-control study. We assess model performance using this case-control data set and

---

**Corresponding author:** Bernard Rosner, 181 Longwood Ave., Boston, MA, 02115, Fax: 617-525-8002, Phone: 617-525-0882, stbar@channing.harvard.edu.

Conflict of interest

No conflicts are present for any author.

evaluate 10-year absolute breast cancer risk. The prospective MMHS validation dataset includes 21.8% of women age <50, and 434 incident cases identified over 10 years of follow-up.

**Results:** In the NHS, MD has the highest odds ratio (OR) for 10-year risk prediction: OR<sub>per SD</sub> = 1.48 (95% CI 1.31 – 1.68), followed by PRS, OR<sub>per SD</sub> = 1.37 (95% CI 1.21 – 1.55) and QS, OR<sub>per SD</sub> = 1.25 (95% CI 1.11 – 1.41). In MMHS, the AUC adjusted for age+MD+QS 0.650; for age+MD+QS+PRS 0.687, and the NRI was 6% in cases and 16% in controls.

**Conclusion:** A simplified assessment of QS, MD and PRS performs consistently to discriminate those at high 10-year breast cancer risk.

**Impact:** This simplified model provides accurate estimation of 10-year risk of invasive breast cancer that can be used in a clinic setting to identify women who may benefit from chemopreventive intervention.

### Keywords

breast cancer; mammographic density; polygenic risk score; validation

---

### Introduction

Breast cancer risk prediction tools are needed to more appropriately stratify risk for women undergoing routine mammographic screening. This demands tools that are easy to use, do not disrupt routine clinical practice, and reflect current science in understanding contributions of factors driving breast cancer risk. Studies show that measures of mammographic breast density (MD), polygenic risk score (PRS), and modifiable and non-modifiable “life course risk factors” are independently related to risk for breast cancer (1,2).

Over the past decade our models were built from life course risk factors that include reproductive factors, adiposity at different points in life, age and type of menopause, menopausal hormone therapy, family history of breast cancer, personal history of benign breast disease, and alcohol intake (3–6). Investigators have added to these models to improve discrimination often measured as the area under the ROC curve (AUC); and include MD (3,7–10); endogenous hormones (11); and/or PRS using a variable number of SNPs (12–18). While combinations of these factors have been evaluated (e.g., lifestyle factors and PRS or lifestyle factors and MD), models that combine measures from all three components (life course risk factors, MD, and PRS) (19–21) show the highest AUC values at 0.71 (9,22). However, comprehensive models present a greater patient burden and simpler models are needed for routine clinic use.

Following fundamental principles of model creation and validation (23,24) we evaluate a simplified model of lifecourse variables reduced from the validated Rosner-Colditz model (25–27) to achieve a simpler parsimonious model that retained performance characteristics of the validated model. We then add MD and PRS, to assess the combined predictive performance in the age range from 45 to 74. We use 10-year absolute risk as the prediction outcome consistent with prediction models for cardiovascular disease and the Tyrer-Cuzick model, used widely in the UK clinical setting (28,29) and ASCO guidelines (30). We report validation using prospective data from the Mayo Mammography Health Study.

## Materials and Methods

We summarize the steps to refine a simplified questionnaire-based score (based on established lifestyle and reproductive factors); then addition of MD and PRS from a case-control data set, drawn from the Nurses' Health Study (NHS) cohort. The study cohort was established in 1976 and included 121,700 female registered nurses aged 30–55 years (31). Questionnaires were mailed to women biennially to collect information on breast cancer risk factors, including age at menarche, age at first birth, parity, family history of breast cancer, height, weight, menopausal status, age at menopause, and hormone therapy use. Alcohol consumption was assessed using a validated semi-quantitative food frequency questionnaire. Women included in these analyses are either pre or postmenopausal and age 45–74 and free from cancer (other than nonmelanoma skin cancer) at the beginning of follow-up in 1990. During 10 years of follow-up time through June 2000, 2799 cases of invasive breast cancer were diagnosed during 770,679 person years.

The study protocol was conducted in accordance with the Declaration of Helsinki and approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. The completion and mailed return of the self-administered questionnaire were considered to imply informed consent.

### Questionnaire-based score (QS)

Risk factors were defined in 1990 without updates during follow-up. Breast cancer risk variables are listed in Supplementary Table S1. We built on the model used to estimate population attributable risk for breast cancer from the NHS (32) and further reduced the pregnancy variables to only include age at first birth, removing parity, and age at menarche. Age and menopausal status were treated as combined indicator variables to account for the rate of rising incidence per year in premenopausal women, and the slower rise in incidence per year among postmenopausal women (4–6). Personal history of benign breast disease confirmed by breast biopsy and family history of breast cancer in a first degree relative were each included. Hormone therapy was classified as past and never combined (reference group) with separate indicators for current form of therapy (estrogen alone; estrogen plus progestin; progesterone alone; and other). Height, and current Body Mass Index (BMI) were modeled as continuous variables and current alcohol as a set of indicators to better capture the increase in risk with higher intake. Risk factors in this model were assessed at baseline (1990) and not updated. The questionnaire score (QS) was developed from a Cox regression model based on these risk factors to predict 10-year risk of invasive breast cancer (1990–2000) and is given by:

$$L = \sum_{i=1}^{10} \beta_i x_i,$$

where  $x_i$  is the  $i$ th risk factor and  $\beta_i$  is the corresponding regression coefficient (Table 2).

## Mammographic breast density (MD)

In the NHS, MD is only available in a nested case-control data set of women who provided blood samples in 1989–1990. That data set includes assessment of MD using Cumulus software to measure percent mammographic density (33,34) with reproducibility between readings of 0.90 (35). It comprises 533 cases of breast cancer and 633 controls matched on age and month of blood draw. We assessed MD for mammograms retrieved from the participants as close to blood draw (1989–1990) for DNA assessment as possible, and use this as a continuous variable (33). In a sensitivity analysis we use the BIRADS score.

This study also includes women with GWAS data. An externally derived polygenic risk score (PRS) was estimated using GWAS data and consists of 77 SNPs (36). The PRS was defined as the sum of the number of risk alleles across 77 SNPs weighted by the independently-estimated effect size of each SNP.

## Analysis

We assess performance over 10 years of follow-up from the baseline assessment and risk classification using the nested case-control data set. An unconditional logistic regression was run on the nested case-control data set of the form:

$$\text{Logit}(p) = \alpha + \beta_1(QS) + \beta_2(MD) + \beta_3(\text{PRS})$$

where  $p$  = probability of breast cancer

The summary risk score is given by

$$S = \beta_1 QS + \beta_2 MD + \beta_3 \text{PRS}$$

To estimate 10-year risk of breast cancer, we follow the approach of Gail to calibrate individual risk to the SEER population breast cancer rates (37), as applied to calibrating the breast cancer incidence model in the California Teachers Study ((25), Supplementary Materials and Methods). After generating individual 10-year risk estimates, we stratify women using the UK guideline categories: below average (<2%), average (2% to <3%), above average (3 to <5%), moderately increased (5% to < 8%), and high ( 8%) (10,28,38).

## Validation study

The prospective Mayo Mammography Health Study (MMHS) enrolled participants undergoing screening mammography who were residents of Minnesota, Iowa or Wisconsin, from October 2003 through September 2006 at the Mayo Clinic in Rochester, MN (39). The baseline questionnaire included risk factor data. Annual follow-up used the Mayo and tri-state tumor registries and mail and/or phone contacts for those who moved outside this region. This study includes incident invasive breast cancer cases through 10 years of follow-up, and an at-risk cohort representing approximately 10% of the overall cohort, 21.8% aged <50 at enrollment. MD was assessed by an expert reader using the Cumulus program. We have previously shown comparable estimates from the NHS and MMHS studies (40). PRS was also generated from the 77 published breast cancer GWAS SNPs (2). Cases and controls

for analysis were limited to 434 incident cases and 898 controls with data on QS, MD, and PRS.

### Validation analysis

For each woman in MMHS, we generated a questionnaire score (QS) using the beta coefficients from the NHS (Table 2) and the woman's own risk factor status. We then summarize QS, MD, and PRS for cases and controls in the independent MMHS data (Table 4). To assess performance in the independent MMHS data we calculate median estimated 10-year risk in each of the 5 risk groups in NHS then run a logistic regression of case/control status on median risk in the MMHS data. This Wald Chi-square provides a test for trend on 1 df. We compare model A age+QS+MD vs model B age+QS+MD+PRS using AUC and NRI (41).

### Results

In Table 1 we describe the prevalence of risk factors in the NHS included in the QS in 1990 and the number of incident invasive breast cancer cases from 1990–2000. Distributions reflect patterns for established breast cancer risk factors. Supplementary Table S2 has the baseline prevalence of risk factors for the nested case-control data set.

Table 2 summarizes regression coefficients, hazard ratios and 95 percent confidence intervals (95% CI) for the variables included in the QS for 10-year risk of breast cancer. We note that among premenopausal women incidence rises with age. Among postmenopausal women who are 45 to 49 years of age, risk is significantly lower than for their comparable aged premenopausal counterparts (HR=0.28, 95% CI 0.16 to 0.50). For postmenopausal women, risk continues to rise through age 74. At any given age for postmenopausal women, longer duration of menopause corresponds to earlier age at menopause and lower risk, HR=0.85 (0.81 to 0.89) per 5 years. Other risk factors are consistent with prior studies.

We used the coefficients from the full model (Table 2) to estimate the QS and then ran a logistic regression in the nested case-control data set of breast cancer risk on QS, MD, and PRS (33). The odds ratio (OR) per SD is presented (Table 3). We see in the mutually adjusted model that MD has the highest OR:  $OR_{\text{per SD}} = 1.48$  (95% CI 1.31 – 1.68) followed by PRS,  $OR_{\text{per SD}} = 1.37$  (95% CI 1.21 – 1.55) and QS,  $OR_{\text{per SD}} = 1.25$  (95% CI 1.11 – 1.41). These component measures also have low correlations with each other; Spearman correlation ( $r_s$ ): (QS vs. MD, 0.054; QS vs. PRS, 0.024; MD vs. PRS, 0.085), consistent with similarity between marginal and mutually-adjusted ORs. Furthermore, there was no significant interaction among QS, MD, and PRS in their effects on incident breast cancer.

We assessed performance of the model both in the NHS and MMHS. (see Table 4). Cases have significantly higher mean scores for 10-year risk than controls both for each component and the overall risk score. The AUC for the overall model adjusted for age was 0.523, for age + MD 0.629, for age + MD + QS 0.637 and for age+MD+QS+PRS 0.658. There was no significant interaction between menopausal status and the components of the overall risk score. Repeating analysis using BIRADS for breast density gave comparable AUC values (see supplementary Table S3).

We next assessed the distribution of cases and controls according to the 10-year estimated risk of breast cancer (Table 5). Using the UK National Health Service cut points for 10-year breast cancer risk, the very high risk category (8% or higher 10-year risk) includes 9% of the controls and 18% of the cases. The lowest category (10-year risk less than 2%) includes 19% of the controls and 9% of the cases. This represents an OR of 4.0 consistent with the ability of the model to separate the distribution of risk in cases and non-cases when predicting 10-year risk of breast cancer. Figure 1 presents these distributions demonstrating separation of controls to the left and cases to the right.

## Validation

From MMHS data we note in table 1 that the women are slightly older with more cases and controls in the 65 to 74 year age range than in the NHS. The MMHS includes more nulliparous women and also a higher proportion with first birth before age 25, and fewer with first birth after age 30 than in the NHS. They also have a higher proportion of women with history of benign breast disease and family history of breast cancer. Fewer women currently use menopausal hormone therapy, more postmenopausal women are obese, and alcohol intake is lower than among the NHS.

Using the beta-coefficients from the NHS, the estimated QS is lower than in the NHS although the mean difference between cases and controls in the MMHS (0.07,  $p=0.006$ ) is similar to the mean difference in the NHS (0.08,  $p<0.0001$ ) (Table 4). MD is lower overall in the MMHS, but the case / control differences are statistically significant, and the PRS is also higher in cases than controls with significant differences overall. The overall AUC is 0.687, comparable to 0.658 in the derivation cohort. Evaluating AUC in MMHS as variables are added, we observed that AUC adjusted for age was 0.595; for age+MD 0.635; for age+MD+QS 0.650; for age+MD+QS+PRS 0.687.

Stratifying women by their overall 10-year estimated breast cancer risk category (table 5) shows comparable cases and controls in the below average 10-year risk category in MMHS (6% of cases, 19% of controls) and NHS (9% of cases and 19% controls). Estimated risk increased significantly with increasing risk category in the MMHS, Wald  $X^2 = 85.83$ ,  $p<0.001$ . Similar distributions of women in the moderately increased and high risk categories are seen across both studies. Figures 1a and 1b demonstrate the separation of the distribution of risk among cases and controls in the two cohorts. The NRI comparing model A age+QS+MD vs model B age+QS+MD+PRS was 6% in cases and 16% in controls (Supplementary Table S4).

## Discussion

A simplified lifestyle assessment of breast cancer risk factors, combined with MD and PRS, performs consistently to discriminate those at high risk of subsequent breast cancer from those who will remain free from breast cancer over 10-years. This combination of three established predictors of breast cancer risk is robust and after controlling for age, it identifies 18% of cases and 9% of controls coming from the high risk category ( $>8\%$  10-year risk). Equally important, only 9% of cases and 19% of controls arise from those at low risk of breast cancer in the next ten years ( $<2\%$  10-year risk). This separation of absolute 10-year



risk for cases and controls holds up in an independent validation data set of over 400 incident invasive breast cancer cases. The AUC is 0.687 in the MMHS validation data set with an NRI of 6% in cases and 16% in controls for Model A vs. Model B. Stratification of risk using a model that integrates questionnaire risk factors, PRS and MD could usefully be adapted to clinical practice to modify frequency and intensity of screening programs.

We note that assessment in colorectal cancer also can perform adequately in the context of simplified assessment by dichotomizing or reducing respondent burden with simpler assessment (42). Truncating assessment complexity does not reduce performance while it increases ease of use (43–47).

MD shows only modest decay over time. The longitudinal correlation of MD over 10 years is estimated at 0.91 (48), and peak percent MD is at ages 35–39 (49). Thus, a one-time assessment is adequate for long-term risk prediction (50). Additionally, germline SNPs are unchanging, equally supporting the need for only a one-time assessment. The number of SNPs in PRS has increased over time from 77 (36) to 313 (15) SNPs with modest change in the AUC. Here we use the lower count as a conservative PRS, but expect similar findings with the expanded PRS, though generating a PRS comprising 300 or more SNPs (15) remains problematic in routine screening clinics.

Risk models continue to pursue dual measures of success – better model performance based on discrimination and easier adaptation to clinical application. We have shown that in an independent validation data set the Rosner-Colditz model of lifecourse questionnaire risk factors outperforms the Gail model (25). Compared to the Rosner-Colditz model, both the Gail and Tyrer-Cuzick models over predict risk in high risk women and under predict risk in low risk women (27). Both models perform less well than the Rosner-Colditz model, and all show better discrimination among women under age 70 (27). Assessing 10-year performance of BOADICEA, BRCAPRO, the Breast Cancer Risk Assessment Tool (BCRAT), and IBIS model, the BCRAT model under predicts risk as does BRCAPRO, and the BOADICEA, BCRAT, and IBIS models over predict in the highest risk subgroup of women, while the models are overall well calibrated with AUC of 0.71 or lower (51). That analysis used data from the Breast Cancer Prospective Family Study Cohort of women without breast cancer at baseline but identified from families over sampled for early onset breast cancer (52).

We use SEER breast cancer incidence data from 1995 for the population estimates of average risk. This corresponds to the mid-point of the 10 year follow-up for the NHS. This anchoring could be readily updated to more recent time periods for routine clinical use.

We, like others (53), see no interaction between MD and PRS attesting to the value of the combined measures to maximize discrimination. The BOADICEA implementation of similar combinations of lifestyle, PRS, and MD gives a broad distribution of risk for the UK population (54). Moving to routine clinical use both to increase screening in high risk women, and reduce screening and surveillance in low-risk women, looks increasingly possible. The NRI of 16% in controls supports adding PRS to the risk model. This is consistent with the growing call for more widespread use of PRS in routine clinical care

(55,56) and continuing refinement to screening recommendations that use risk of disease to modify frequency and modality for screening (57). Saliva DNA can be collected prior to baseline mammogram to generate PRS.

Current US guidelines vary, but in general recommend screening mammography from age 45 (American Cancer Society (58)) or 50 (US Preventive Services Task Force (59)), with either annual or biennial mammography (60). Assessment and screening decisions should include shared decision making as recommended by numerous guidelines. This assumes a reliable risk assessment and shared decision-making tools. These recommendations apply to average risk women, defined by USPSTF to include women with a history of breast cancer in a single family member and women with dense breasts.

In summary, the combination of QS, MD, and PRS predicts 10-year risk of invasive breast cancer. Now validated, this model can be readily implemented in routine clinical services to identify women at higher risk of breast cancer and those who are at low risk and unlikely to benefit from chemoprevention interventions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We would like to thank the participants and staff of the Nurses' Health Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

Financial support

This project was funded by a National Institutes of Health cohort infrastructure Grant (UM1 CA186107), and a program project Grant (P01 CA87969) from the National Cancer Institute, and by a grant (BCRF 20-028) to Dr. Colditz from the Breast Cancer Research Foundation. Mayo Mammography Health Study was funded by NCI R01 CA97396, R01 CA177150 and Mayo Clinic Cancer Center.

The funding agencies had no role in the construction or the writing of this manuscript.

## References

1. Cuzick J, Brentnall A, Dowsett M. SNPs for breast cancer risk assessment. *Oncotarget* 2017;8(59):99211–2 doi 10.18632/oncotarget.22278. [PubMed: 29245890]
2. Vachon CM, Pankratz VS, Scott CG, Haerberle L, Ziv E, Jensen MR, et al. The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst* 2015;107(5) doi 10.1093/jnci/dju397.
3. Rice MS, Tworoger SS, Hankinson SE, Tamimi RM, Eliassen AH, Willett WC, et al. Breast cancer risk prediction: an update to the Rosner-Colditz breast cancer incidence model. *Breast cancer research and treatment* 2017;166(1):227–40 doi 10.1007/s10549-017-4391-5. [PubMed: 28702896]
4. Rosner B, Colditz GA, Willett WC. Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. *American journal of epidemiology* 1994;139(8):819–35 doi 10.1093/oxfordjournals.aje.a117079. [PubMed: 8178795]
5. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. *Journal of the National Cancer Institute* 1996;88(6):359–64 doi 10.1093/jnci/88.6.359. [PubMed: 8609645]



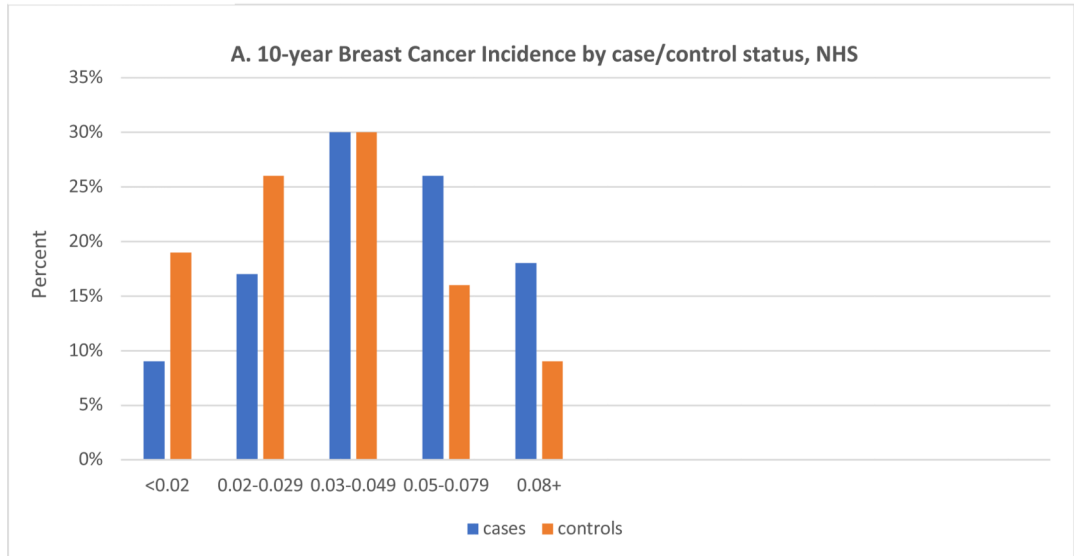
6. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *American journal of epidemiology* 2000;152(10):950–64 doi 10.1093/aje/152.10.950. [PubMed: 11092437]
7. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat* 2005;94(2):115–22 doi 10.1007/s10549-005-5152-4. [PubMed: 16261410]
8. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med* 2008;148(5):337–47 doi 10.7326/0003-4819-148-5-200803040-00004. [PubMed: 18316752]
9. Eriksson M, Czene K, Pawitan Y, Leifland K, Darabi H, Hall P. A clinical model for identifying the short-term risk of breast cancer. *Breast Cancer Res* 2017;19(1):29 doi 10.1186/s13058-017-0820-y. [PubMed: 28288659]
10. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density. *JAMA Oncol* 2018;4(9):e180174 doi 10.1001/jamaoncol.2018.0174. [PubMed: 29621362]
11. Husing A, Fortner RT, Kuhn T, Overvad K, Tjonneland A, Olsen A, et al. Added Value of Serum Hormone Measurements in Risk Prediction Models for Breast Cancer for Women Not Using Exogenous Hormones: Results from the EPIC Cohort. *Clin Cancer Res* 2017;23(15):4181–9 doi 10.1158/1078-0432.CCR-16-3011. [PubMed: 28246273]
12. Shieh Y, Hu D, Ma L, Huntsman S, Gard CC, Leung JWT, et al. Joint relative risks for estrogen receptor-positive breast cancer from a clinical model, polygenic risk score, and sex hormones. *Breast Cancer Res Treat* 2017;166(2):603–12 doi 10.1007/s10549-017-4430-2. [PubMed: 28791495]
13. Maas P, Barrdahl M, Joshi AD, Auer PL, Gaudet MM, Milne RL, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol* 2016;2(10):1295–302 doi 10.1001/jamaoncol.2016.1025. [PubMed: 27228256]
14. van Veen EM, Brentnall AR, Byers H, Harkness EF, Astley SM, Sampson S, et al. Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. *JAMA Oncol* 2018;4(4):476–82 doi 10.1001/jamaoncol.2017.4881. [PubMed: 29346471]
15. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2019;104(1):21–34 doi 10.1016/j.ajhg.2018.11.002. [PubMed: 30554720]
16. Brentnall AR, van Veen EM, Harkness EF, Rafiq S, Byers H, Astley SM, et al. A case-control evaluation of 143 single nucleotide polymorphisms for breast cancer risk stratification with classical factors and mammographic density. *Int J Cancer* 2020;146(8):2122–9 doi 10.1002/ijc.32541. [PubMed: 31251818]
17. Lakeman IMM, Hilbers FS, Rodriguez-Gironde M, Lee A, Vreeswijk MPG, Hollestelle A, et al. Addition of a 161-SNP polygenic risk score to family history-based risk prediction: impact on clinical management in non-BRCA1/2 breast cancer families. *J Med Genet* 2019;56(9):581–9 doi 10.1136/jmedgenet-2019-106072. [PubMed: 31186341]
18. Lall K, Lepamets M, Palover M, Esko T, Metspalu A, Tonisson N, et al. Polygenic prediction of breast cancer: comparison of genetic predictors and implications for risk stratification. *BMC cancer* 2019;19(1):557 doi 10.1186/s12885-019-5783-1. [PubMed: 31182048]
19. Zhang X, Rice M, Tworoger SS, Rosner BA, Eliassen AH, Tamimi RM, et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. *PLoS Med* 2018;15(9):e1002644 doi 10.1371/journal.pmed.1002644. [PubMed: 30180161]
20. Pal Choudhury P, Wilcox AN, Brook MN, Zhang Y, Ahearn T, Orr N, et al. Comparative Validation of Breast Cancer Risk Prediction Models and Projections for Future Risk Stratification. *J Natl Cancer Inst* 2020;112(3):278–85 doi 10.1093/jnci/djz113. [PubMed: 31165158]
21. Evans DGR, Harkness EF, Brentnall AR, van Veen EM, Astley SM, Byers H, et al. Breast cancer pathology and stage are better predicted by risk stratification models that include mammographic

- density and common genetic variants. *Breast Cancer Res Treat* 2019;176(1):141–8 doi 10.1007/s10549-019-05210-2. [PubMed: 30941651]
22. Louro J, Posso M, Hilton Boon M, Roman M, Domingo L, Castells X, et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br J Cancer* 2019;121(1):76–85 doi 10.1038/s41416-019-0476-8. [PubMed: 31114019]
  23. Pencina MJ, Goldstein BA, D’Agostino RB. Prediction Models - Development, Evaluation, and Clinical Application. *N Engl J Med* 2020;382(17):1583–6 doi 10.1056/NEJMp2000589. [PubMed: 32320568]
  24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162(1):55–63 doi 10.7326/M14-0697. [PubMed: 25560714]
  25. Rosner BA, Colditz GA, Hankinson SE, Sullivan-Halley J, Lacey JV Jr., Bernstein L. Validation of Rosner-Colditz breast cancer incidence model using an independent data set, the California Teachers Study. *Breast cancer research and treatment* 2013;142(1):187–202 doi 10.1007/s10549-013-2719-3. [PubMed: 24158759]
  26. Glynn RJ, Colditz GA, Tamimi RM, Chen WY, Hankinson SE, Willett WW, et al. Extensions of the Rosner-Colditz breast cancer prediction model to include older women and type-specific predicted risk. *Breast cancer research and treatment* 2017;165(1):215–23 doi 10.1007/s10549-017-4319-0. [PubMed: 28589369]
  27. Glynn RJ, Colditz GA, Tamimi RM, Chen WY, Hankinson SE, Willett WW, et al. Comparison of Questionnaire-Based Breast Cancer Prediction Models in the Nurses’ Health Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2019;28(7):1187–94 doi 10.1158/1055-9965.EPI-18-1039.
  28. Evans DG, Donnelly LS, Harkness EF, Astley SM, Stavrinou P, Dawe S, et al. Breast cancer risk feedback to women in the UK NHS breast screening population. *Br J Cancer* 2016;114(9):1045–52 doi 10.1038/bjc.2016.56. [PubMed: 27022688]
  29. Evans DG, Brentnall AR, Harvie M, Astley S, Harkness EF, Stavrinou P, et al. Breast cancer risk in a screening cohort of Asian and white British/Irish women from Manchester UK. *BMC Public Health* 2018;18(1):178 doi 10.1186/s12889-018-5090-9. [PubMed: 29370783]
  30. Visvanathan K, Fabian CJ, Bantug E, Brewster AM, Davidson NE, DeCensi A, et al. Use of Endocrine Therapy for Breast Cancer Risk Reduction: ASCO Clinical Practice Guideline Update. *J Clin Oncol* 2019;37(33):3152–65 doi 10.1200/JCO.19.01472. [PubMed: 31479306]
  31. Colditz GA, Hankinson SE. The Nurses’ Health Study: lifestyle and health among women. *Nature reviews Cancer* 2005;5(5):388–96 doi 10.1038/nrc1608. [PubMed: 15864280]
  32. Tamimi RM, Spiegelman D, Smith-Warner SA, Wang M, Pazaris M, Willett WC, et al. Population Attributable Risk of Modifiable and Nonmodifiable Breast Cancer Risk Factors in Postmenopausal Breast Cancer. *Am J Epidemiol* 2016;184(12):884–93 doi 10.1093/aje/kww145. [PubMed: 27923781]
  33. Tamimi RM, Colditz GA, Hankinson SE. Circulating carotenoids, mammographic density, and subsequent risk of breast cancer. *Cancer Res* 2009;69(24):9323–9 doi 10.1158/0008-5472.CAN-09-1018. [PubMed: 19934322]
  34. Byng JW, Boyd NF, Little L, Lockwood G, Fishell E, Jong RA, et al. Symmetry of projection in the quantitative analysis of mammographic images. *Eur J Cancer Prev* 1996;5(5):319–27 doi 10.1097/00008469-199610000-00003. [PubMed: 8972250]
  35. Pettersson A, Hankinson SE, Willett WC, Lagiou P, Trichopoulos D, Tamimi RM. Nondense mammographic area and risk of breast cancer. *Breast Cancer Res* 2011;13(5):R100 doi 10.1186/bcr3041. [PubMed: 22017857]
  36. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 2015;107(5) doi 10.1093/jnci/djv036.
  37. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting Individualized Probabilities of Developing Breast-Cancer for White Females Who Are Being Examined

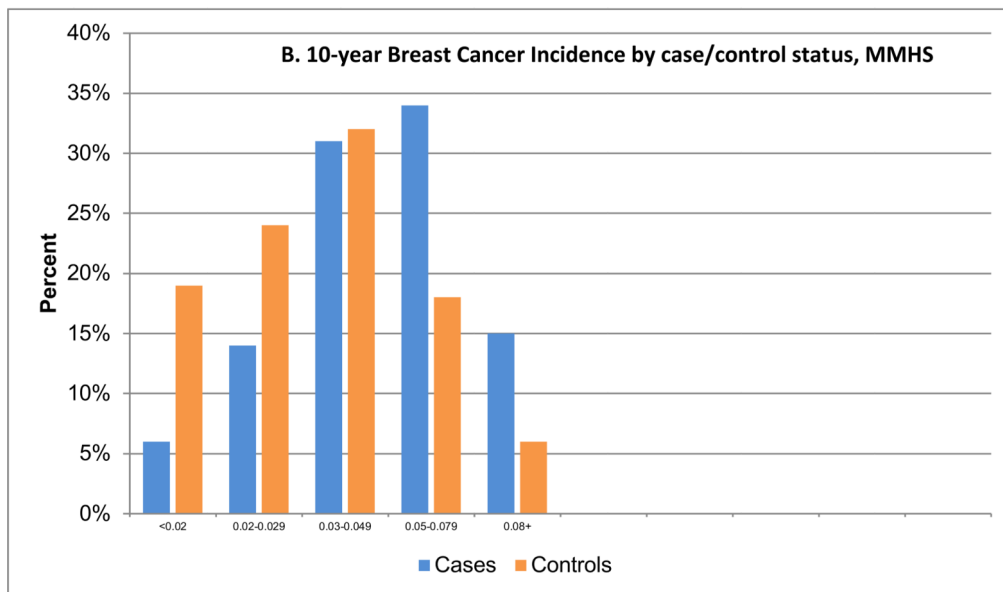
- Annually. *Journal of the National Cancer Institute* 1989;81(24):1879–86 doi DOI 10.1093/jnci/81.24.1879. [PubMed: 2593165]
38. Brentnall AR, Harkness EF, Astley SM, Donnelly LS, Stavrinou P, Sampson S, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res* 2015;17(1):147 doi 10.1186/s13058-015-0653-5. [PubMed: 26627479]
  39. Olson JE, Sellers TA, Scott CG, Schueler BA, Brandt KR, Serie DJ, et al. The influence of mammogram acquisition on the mammographic density and breast cancer association in the Mayo Mammography Health Study cohort. *Breast Cancer Res* 2012;14(6):R147 doi 10.1186/bcr3357. [PubMed: 23152984]
  40. Bertrand KA, Tamimi RM, Scott CG, Jensen MR, Pankratz V, Visscher D, et al. Mammographic density and risk of breast cancer by age and tumor characteristics. *Breast Cancer Res* 2013;15(6):R104 doi 10.1186/bcr3570. [PubMed: 24188089]
  41. Pencina MJ, D'Agostino RB Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;30(1):11–21 doi 10.1002/sim.4085. [PubMed: 21204120]
  42. Liu Y, Colditz GA, Rosner BA, Dart H, Wei E, Waters EA. Comparison of Performance Between a Short Categorized Lifestyle Exposure-based Colon Cancer Risk Prediction Tool and a Model Using Continuous Measures. *Cancer prevention research* 2018;11(12):841–8 doi 10.1158/1940-6207.CAPR-18-0196. [PubMed: 30446519]
  43. Rifas-Shiman SL, Willett WC, Lobb R, Kotch J, Dart C, Gillman MW. PrimeScreen, a brief dietary screening tool: reproducibility and comparability with both a longer food frequency questionnaire and biomarkers. *Public Health Nutr* 2001;4(2):249–54 doi 10.1079/phn200061. [PubMed: 11299098]
  44. Usher-Smith JA, Walter FM, Emery JD, Win AK, Griffin SJ. Risk Prediction Models for Colorectal Cancer: A Systematic Review. *Cancer Prev Res (Phila)* 2016;9(1):13–26 doi 10.1158/1940-6207.CAPR-15-0274. [PubMed: 26464100]
  45. Smith T, Muller DC, Moons KGM, Cross AJ, Johansson M, Ferrari P, et al. Comparison of prognostic models to predict the occurrence of colorectal cancer in asymptomatic individuals: a systematic literature review and external validation in the EPIC and UK Biobank prospective cohort studies. *Gut* 2019;68(4):672–83 doi 10.1136/gutjnl-2017-315730. [PubMed: 29615487]
  46. Galesic M BM. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly* 2009;73(2):349–60.
  47. Ovaskainen ML, Paturi M, Reinivuo H, Hannila ML, Sinkko H, Lehtisalo J, et al. Accuracy in the estimation of food servings against the portions in food photographs. *European Journal of Clinical Nutrition* 2008;62(5):674–81 doi 10.1038/sj.ejcn.1602758. [PubMed: 17440523]
  48. Krishnan K, Baglietto L, Stone J, Simpson JA, Severi G, Evans CF, et al. Longitudinal Study of Mammographic Density Measures That Predict Breast Cancer Risk. *Cancer Epidemiol Biomarkers Prev* 2017;26(4):651–60 doi 10.1158/1055-9965.EPI-16-0499. [PubMed: 28062399]
  49. Burton A, Maskarinec G, Perez-Gomez B, Vachon C, Miao H, Lajous M, et al. Mammographic density and ageing: A collaborative pooled analysis of cross-sectional data from 22 countries worldwide. *PLoS Med* 2017;14(6):e1002335 doi 10.1371/journal.pmed.1002335. [PubMed: 28666001]
  50. Yaghjian L, Colditz GA, Rosner B, Tamimi RM. Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to the time since the mammogram. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2013;22(6):1110–7 doi 10.1158/1055-9965.EPI-13-0169.
  51. Terry MB, Liao Y, Whittemore AS, Leco N, Buchsbaum R, Zeinomar N, et al. 10-year performance of four models of breast cancer risk: a validation study. *Lancet Oncol* 2019;20(4):504–17 doi 10.1016/S1470-2045(18)30902-1. [PubMed: 30799262]
  52. Terry MB, Phillips KA, Daly MB, John EM, Andrulis IL, Buys SS, et al. Cohort Profile: The Breast Cancer Prospective Family Study Cohort (ProF-SC). *Int J Epidemiol* 2016;45(3):683–92 doi 10.1093/ije/dyv118. [PubMed: 26174520]

53. Vachon CM, Scott CG, Tamimi RM, Thompson DJ, Fasching PA, Stone J, et al. Joint association of mammographic density adjusted for age and body mass index and polygenic risk score with breast cancer risk. *Breast Cancer Res* 2019;21(1):68 doi 10.1186/s13058-019-1138-8. [PubMed: 31118087]
54. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med* 2019;21(8):1708–18 doi 10.1038/s41436-018-0406-9. [PubMed: 30643217]
55. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;50(9):1219–24 doi 10.1038/s41588-018-0183-z. [PubMed: 30104762]
56. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet* 2019;28(R2):R133–R42 doi 10.1093/hmg/ddz187. [PubMed: 31363735]
57. Monticciolo DL, Newell MS, Moy L, Niell B, Monsees B, Sickles EA. Breast Cancer Screening in Women at Higher-Than-Average Risk: Recommendations From the ACR. *J Am Coll Radiol* 2018;15(3 Pt A):408–14 doi 10.1016/j.jacr.2017.11.034. [PubMed: 29371086]
58. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA* 2015;314(15):1599–614 doi 10.1001/jama.2015.12783. [PubMed: 26501536]
59. Siu AL, Force USPST. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med* 2016;164(4):279–96 doi 10.7326/M15-2886. [PubMed: 26757170]
60. Qaseem A, Lin JS, Mustafa RA, Horwitch CA, Wilt TJ, Clinical Guidelines Committee of the American College of P. Screening for Breast Cancer in Average-Risk Women: A Guidance Statement From the American College of Physicians. *Ann Intern Med* 2019;170(8):547–60 doi 10.7326/M18-2147. [PubMed: 30959525]

a



b



**Figure 1.**

shows estimated 10-year incidence of breast cancer by case/control status

Figure 1a contains a histogram of estimated 10-year breast cancer incidence in the NHS, separately for cases (blue) and controls (red).

Figure 1b contains a histogram of estimated 10-year breast cancer incidence in the MMHS, separately for cases (blue) and controls (red).

Distribution of Questionnaire Risk Factors in Nurses' Health Study data for women age  $\geq 45$  in 1990 for women who developed breast cancer over 10 years of follow-up and those who remained free from cancer, and for Mayo Mammography Health Study

**Table 1:**

Variable categories	Nurses' Health Study		Mayo Mammography Health Study	
	Cases N(%)	Controls N(%)	Cases N(%)	Controls N(%)
<b>N</b>	2799	75557	438	898
<b>Age, menopausal status, N(%)</b>				
age 45–49	368(13.2%)	13514(17.9%)	51 (11.6%)	196 (21.8%)
age 50–54	557(19.9%)	16877(22.3%)	78 (17.8%)	181 (20.2%)
age 55–59	634(22.7%)	17010(22.5%)	71 (16.2%)	155 (17.3%)
age 60–64	683(24.4%)	15540(20.6%)	105 (24.0%)	127 (14.1%)
age 65–74	557(19.9%)	12616(16.7%)	133 (30.4%)	239 (26.6%)
<b>Menopausal status, N(%)</b>				
Premenopausal	502(17.9%)	15551(20.6%)	74 (16.9%)	222 (24.7%)
Postmenopausal	2297(82.1%)	60004(79.4%)	364 (83.1%)	676 (75.3%)
<b>Duration of post-menopause, years, mean(sd)<sup>a</sup></b>	10.1(6.2)	10.1(6.5)	14.5 (8.7)	13.4 (9.0)
<b>Pregnancy history</b>				
Nulliparous	204(7.3%)	4254(5.6%)	51 (11.6%)	116 (12.9%)
age 1st birth 20–24	1199(42.8%)	36977(48.9%)	255 (57.5%)	508 (56.5%)
age 1st birth 25–29	1071(38.3%)	27365(36.2%)	92 (21.0%)	183 (20.4%)
age 1st birth 30+	325(11.6%)	6959(9.2%)	32 (7.3%)	80 (8.9%)
<b>BBD (Biopsy confirmed), N(%)</b>	717(25.6%)	14551(19.3%)	158 (35.8%)	197 (21.9%)
<b>Family history of breast cancer, N(%)</b>	422(15.1%)	7672(10.2%)	119 (27.1%)	174 (19.4%)
<b>Current menopausal hormone therapy use, N(%)<sup>b</sup></b>				
Not current (past and never users)	1779(63.6%)	52612(69.6%)	331 (75.6%)	686 (76.4%)
Estrogen only	466(16.7%)	11278(14.9%)	64 (14.6%)	145 (16.5%)
Estrogen & Progestin	399(14.3%)	7050(9.3%)	30 (6.8%)	42 (4.7%)
Progesterone alone	18(0.6%)	487(0.6%)	3 (0.7%)	3 (0.3%)
Other <sup>c</sup>	137(4.9%)	4130(5.5%)	10 (2.3%)	35 (3.8%)
<b>Current BMI, pre-menopause, N(%)</b>				



Variable categories	Nurses' Health Study		Mayo Mammography Health Study	
	Cases N(%)	Controls N(%)	Cases N(%)	Controls N(%)
<25	286(57.0%)	8475(54.5%)	34(45.9%)	86(38.7%)
25–29.9	147(29.3%)	4390(28.2%)	24(32.4%)	69(31.1%)
>=30	69(13.8%)	2687(17.3%)	16(21.6%)	67(30.2%)
mean(sd)	25.2(4.4)	25.8(5.1)	27.1(6.8)	28.2(6.6)
<b>Current BMI, post-menopause, N(%)</b>				
<25	1143(49.8%)	30903(51.5%)	96(26.4%)	233(34.5%)
25–29.9	735(32.0%)	18988(31.6%)	125(34.3%)	256(37.9%)
>=30	419(18.2%)	10114(16.9%)	143(39.3%)	187(27.7%)
mean(sd)	26.0(4.9)	25.8(4.9)	29.2(6.1)	27.9(5.9)
<b>Height, inches, mean(sd)</b>	64.7(2.4)	64.5(2.4)	64.6(2.5)	64.5(2.4)
<b>Alcohol, N(%)</b>				
None	1033(36.9%)	29777(39.4%)	230(52.5%)	479(53.3%)
<11 gm/day	1257(44.9%)	33288(44.1%)	192(43.8%)	374(41.6%)
11–21.9 gm/day	313(11.2%)	7929(10.5%)	16(3.7%)	43(1.1%)
>22 gm/day	196(7.0%)	4561(6.0%)	0	2(0.2%)

<sup>g</sup>Hormone therapy status unknown or missing use were deleted

<sup>h</sup>Vaginal estrogen or vaginal progesterone or Other Brands, mixed use or Current user no brand given

<sup>c</sup>among post-menopausal women

Questionnaire score regression coefficients in relation to 10-year risk of breast cancer based on NHS data for women age  $\geq 45$  in 1990 from 1990 to 2000, 2799 cases, 770659 person years

**Table 2:**

Variable categories	unit	beta	SE	P-value	HR	95% CI
<b>age, menopausal status</b>						
age 45–49, pre(ref)	1				1.0	
age 50–59, pre	1	0.242	0.091	0.01	1.27	(1.07,1.52)
age 45–49, post	1	-1.277	0.279	<0.0001	0.28	(0.16,0.50)
age 50–54, post	1	-0.854	0.273	0.002	0.43	(0.25,0.73)
age 55–59, post	1	-0.540	0.273	0.05	0.58	(0.34,1.00)
age 60–64	1	-0.188	0.276	0.50	0.83	(0.48,1.42)
age 65–74	1	-0.024	0.282	0.93	0.98	(0.56,1.70)
<b>duration of post-menopause (per 5 year)</b>	5	-0.162	0.026	<0.0001	0.85	(0.81,0.89)
<b>Pregnancy History</b>						
Nulliparous	1	0.376	0.077	<0.0001	1.46	(1.25,1.69)
age 1st birth 20–24 (ref)	1				1.0	
age 1st birth 25–29	1	0.151	0.043	0.0004	1.16	(1.07,1.27)
age 1st birth 30+	1	0.286	0.063	<0.0001	1.33	(1.18,1.51)
<b>BBD (Biopsy confirmed)</b>	1	0.318	0.044	<0.0001	1.37	(1.26,1.50)
<b>Family history BRCN</b>	1	0.292	0.060	<0.0001	1.34	(1.19,1.51)
<b>Current PMH use <sup>a</sup></b>						
No(ref)	1				1.0	
Estrogen alone	1	0.292	0.055	<0.0001	1.34	(1.20,1.49)
E&P	1	0.573	0.060	<0.0001	1.77	(1.58,1.99)
Progesterone alone	1	0.199	0.238	0.404	1.22	(0.77,1.95)
Other <sup>b</sup>	1	0.041	0.090	0.649	1.04	(0.87,1.24)
<b>Current BMI(per 8 kg/m<sup>2</sup>) <sup>c</sup></b>						
premenopausal	8	-0.122	0.075	0.101	0.89	(0.77,1.02)
postmenopausal	8	0.136	0.034	<0.0001	1.15	(1.07,1.23)
<b>Height (per 6 inches)</b>	6	0.198	0.047	<0.0001	1.22	(1.11,1.34)
<b>Current Alcohol</b>						

Variable categories	unit	beta	SE	P-value	HR	95% CI
None(ref)	1				1.0	
<11 gm/day	1	0.085	0.042	0.05	1.09	(1.00,1.18)
11–21.9 gm/day	1	0.105	0.07	0.12	1.11	(0.98,1.26)
>22 gm/day	1	0.195	0.079	0.01	1.22	(1.04,1.42)

<sup>a</sup>PMH status unknown or missing hormone use were deleted

<sup>b</sup>Vaginal estrogen or vaginal progesterone or Other Brands, mixed use or Current user no brand given

<sup>c</sup>centered at 25 kg/m<sup>2</sup>

Distribution of questionnaire score, percent density of mammogram, and polygenic risk score, beta coefficient and Odds Ratio for 10-year risk of breast cancer from 1990 to 2000, Nurses' Health Study, marginal and mutually adjusted results<sup>a</sup>

**Table 3.**

Variable	SD	Beta	SE	P-value	OR <sup>b</sup>	95% CI
<b>Marginal results</b>						
Questionnaire score	0.3682	0.6573	0.1631	<0.0001	1.27	1.13 1.43
Percent mammographic density	19.2902	0.0219	0.0032	<0.0001	1.53	1.35 1.72
Polygenic risk score	0.9933	0.3432	0.0696	<0.0001	1.41	1.25 1.59
<b>Mutually adjusted results</b>						
Questionnaire score	0.3682	0.6095	0.1682	0.0003	1.25	1.11 1.41
Percent mammographic density	19.2902	0.0204	0.00325	<0.0001	1.48	1.31 1.68
Polygenic risk score	0.9933	0.3154	0.0631	<0.0001	1.37	1.21 1.55

<sup>a</sup>Based on 533 cases and 633 controls.

<sup>b</sup>OR per SD.

**Table 4:** Mean Risk Scores for cases and controls and age adjusted AUC's for 10-year risk prediction applied to case control data among women age >= 45 in 1990 from the Nurses' Health Study, and from the Mayo Mammography Health Study validation data

A. Nurses' Health Study						
Category	Cases mean±sd	N	Controls mean±sd	N	p-value <sup>a</sup>	AUC C-Statistic <sup>b</sup>
Overall	2.35±0.38	533	2.26±0.36	633	<0.0001	0.658
Percent mammographic density	32.28±19.72		24.35±18.16		<0.0001	
Polygenic risk score	0.15±0.98		-0.18±0.98		<0.0001	
Overall risk score	2.13±0.58		1.82±0.55		<0.0001	
Age only						0.523
Age + percent density						0.629
Age + percent density + questionnaire score						0.637
Age + percent density + questionnaire score + polygenic risk score						0.658
B. Mayo Mammography Health Study						
Category	Cases mean±sd	N	Controls mean±sd	N	p-value*	AUC C-Statistic**
Overall	1.97±0.46	438	1.90±0.49	898	0.0063	0.687
Percent mammographic density	18.36±12.41		15.45±11.17		<0.0001	
Polygenic risk score	0.25±0.97		-0.15±0.98		<0.0001	
Overall risk score	1.62±0.51		1.39±0.51		<0.0001	
Age only						0.595
Age + percent density						0.635
Age + percent density + questionnaire score						0.650
Age + percent density + questionnaire score + polygenic risk score						0.687

Note:

<sup>a</sup> based on t test

based on logistic regression including strata of age and then additional risk factors as noted

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 5:**

Distribution of expected 10-year breast cancer risk by case control status

10-year expected incidence of breast cancer	Nurses' Health Study		Mayo Mammography Health Study		Model performance in MMHS cohort	
	case(N=533)	control(N=633)	Cases (N=438)	Controls(N=898)		
<0.02 (below average)	48(9%)	119(19%)	28 (6%)	172 (19%)		
0.02–0.029 (average)	91(17%)	162(26%)	61 (14%)	220 (24%)		
0.03–0.049 (above average)	161(30%)	188(30%)	134 (31%)	290 (32%)		
0.05–0.079 (moderately increased)	137(26%)	104(16%)	148 (34%)	159 (18%)		$\chi^2 = 85.83,$
0.08+ (high)	96(18%)	60(9%)	67 (15%)	57 (6%)		$p < 0.001$