Published in final edited form as:

Breast Cancer Res Treat. 2017 November ; 166(1): 227–240. doi:10.1007/s10549-017-4391-5.

Breast cancer risk prediction: An update to the Rosner-Colditz breast cancer incidence model

Megan S Rice¹, Shelley S Tworoger^{2,3}, Susan E Hankinson^{2,3,4}, Rulla M Tamimi^{2,3}, A Heather Eliassen^{2,3}, Walter C Willett^{2,3,5}, Graham Colditz⁶, and Bernard Rosner¹

¹Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

²Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

³Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁴Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, USA

⁵Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁶Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA

Abstract

Purpose—To update and expand the Rosner-Colditz breast cancer incidence model by evaluating the contributions of more recently identified risk factors as well as predicted percent mammographic density (MD) to breast cancer risk.

Methods—Using data from the Nurses' Health Study (NHS) and NHSII, we added adolescent somatotype (9 unit scale), vegetable intake (servings/day), breastfeeding (months), physical activity (MET-hrs/week), and predicted percent MD to the Rosner-Colditz model to determine whether these variables improved model discrimination. We evaluated all invasive as well as ER+/PR+, ER+/PR–, and ER–/PR– breast cancer.

Results—In the NHS/NHSII, we accrued over 5,200 cases of invasive breast cancer over more than 20 years of follow-up with complete data on the risk factors. Adolescent somatotype and predicted percent MD significantly improved the original Rosner-Colditz model for all invasive breast cancer (change in age-adjusted AUC=0.020, $p<0.001$). The relative risk (RR) of invasive breast cancer for a 4-unit increase in adolescent somatotype was 0.62 (95% CI: 0.56, 0.70) whereas the RR for a 20-unit increase in predicted percent MD was 1.32 (95% CI: 1.28, 1.36). Adolescent somatotype and predicted percent MD also significantly improved the ER+/PR+ model (change in age-adjusted AUC=0.020, $p<0.001$) as well as the ER+/PR– model (change in age-adjusted AUC=0.012, $p=0.007$). Adolescent somatotype, predicted percent MD, breastfeeding, and

Corresponding author: Megan S Rice, ScD, Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, 55 Fruit Street, Bartlett 9, Boston, MA 02114, Tel: (617) 726-8502, mrice1@mgh.harvard.edu.

Conflict of interest: None of the authors has a competing interest to declare.

vegetable intake improved the ER–/PR– model (change in AUC=0.031, $p<0.0001$). The RR of ER–/PR– disease for 5 vegetable servings/day increase was 0.83 (95%CI: 0.70, 0.99), while the RR for every 12 months of breastfeeding was 0.88 (95%CI: 0.77, 1.01). Physical activity did not improve risk classification in any model.

Conclusion—Adolescent somatotype and predicted percent MD significantly improved breast cancer risk classification using the Rosner-Colditz model. Further, risk factors specific to ER– disease, such as breastfeeding and vegetable intake, may also help improve risk prediction of this aggressive subtype.

Keywords

breast cancer; risk prediction; epidemiology

Introduction

In the years since the development and expansion of the Rosner-Colditz breast cancer incidence model,[1, 2] several additional risk factors have been identified in the Nurses' Health Studies (NHS and NHSII) as well as other studies. Specifically, heavier somatotype in adolescence has been inversely associated with breast cancer risk in adulthood, independent of adult body size [3–8]. Greater physical activity has been associated with a lower risk of breast cancer.[9–14] While vegetable intake has not been associated with overall breast cancer risk, higher intakes have been associated with a lower risk of estrogen receptor (ER) negative tumors [15, 16], including in a pooled analysis of 20 cohort studies. [17] Longer duration of breastfeeding has also been associated with lower risk of ER– tumors.[18, 19] Whether these risk factors improve breast cancer risk prediction above and beyond the risk factors in the current Rosner-Colditz model is unknown. Percent mammographic density (MD), which reflects the proportion of the breast which is dense (radiographically light) on a mammogram, is consistently one of the strongest independent predictors of breast cancer risk.[20] Recent research, utilizing the Tyrer-Cuzick and Gail models, suggests that the incorporation of percent MD may improve risk prediction.[21, 22] While mammograms are available for only a subset of the NHS/NHSII, we previously developed a model to predict percent MD for the majority of women in the cohorts [Rice et al., *Cancer Causes and Control*, in press]. Therefore, we examined whether adolescent somatotype, physical activity, breastfeeding, vegetable intake, and predicted percent MD improved the discriminatory ability of the Rosner-Colditz model for all invasive breast cancer as well as for ER+/PR+, ER+, PR–, and ER–/PR– disease.

Materials and Methods

Study population

The NHS and NHSII have been described previously.[1, 2] Briefly, 121,700 female registered nurses aged 30–55 years of age from 11 U.S. states completed an initial questionnaire in 1976 forming the NHS cohort. The NHSII began in 1989 when 116,429 female registered nurses, aged 25 to 42, residing in 14 U.S. states completed a baseline questionnaire. Both cohorts are followed by biennially mailed questionnaires to collect

information on exposures and covariates as well as incident diseases. The study was approved by the Committee on the Use of Human Subjects in Research at the Brigham and Women's Hospital.

Selected breast cancer risk factors

A number of risk factors are included in the previously published Rosner-Colditz risk model[1, 2]. On the baseline questionnaires, participants reported their height and age at menarche. Current weight and personal history of benign breast disease (BBD) were queried on all biennial questionnaires. Beginning in 1982 (NHS) and 1989 (NHSII), women were asked whether their BBD was confirmed by biopsy. Weight at age 18 was asked in 1980 (NHS) and in 1989 (NHSII). Current body mass index (BMI) and BMI at age 18 were estimated as $\text{weight(kg)/height(m)}^2$. Participants reported whether they had a first-degree relative (mother or sister) with a diagnosis of breast cancer at baseline, in 1982 (NHS), and then every 4 years beginning in 1988 (NHS) or 1997 (NHSII). In the NHS, parity and age at first birth were queried on the baseline questionnaire. In 1978, participants updated information on parity. Data on parity were updated on biennial questionnaires until 1984 and again in 1996. In NHSII, parity and ages at all births were asked on the baseline as well as subsequent biennial questionnaires. Birth index, a summary variable representing both the number and spacing of pregnancies, was estimated as the sum of the total years from each birth to a woman's current age (or age at menopause if postmenopausal) with nulliparous women receiving a value of 0. In the NHS, current alcohol intake was first asked in 1980 and subsequently in 1984, 1986, and every four years afterwards. In the NHSII, current alcohol intake was asked at baseline, in 1991 and every four years afterwards. In 1988 (NHS) and 1989 (NHSII), participants responded to questions on alcohol intake at various ages prior to enrollment (e.g., ages 18–22). On baseline and biennial questionnaires, women were asked if their periods had ceased and whether this was natural or due to bilateral oophorectomy and/or hysterectomy. Women were asked about their postmenopausal hormone therapy (HT) use on the baseline and biennial questionnaires, including type of HT (estrogen only [E], estrogen and progestin [E+P], other) starting in 1978 (NHS) or 1989 (NHSII).

Candidate risk factors

For risk factors that we evaluated for inclusion into the updated model, in 1988 (NHS) and 1989 (NHSII), participants were asked to report their body size at ages 10 and 20 using a 9-level figure (1=leanest figure, 9=heaviest figure).[23] Adolescent somatotype was estimated as the average of somatotypes at ages 10 and 20. Breastfeeding history was asked in 1986 in NHS and in 1993 and 1997 in NHSII. Physical activity was assessed in 1986 (NHS) and 1989 (NHSII) and every 2–6 years afterwards. Participants reported the average time per week over the prior year spent doing various activities (e.g., walking, running). Metabolic equivalent task (MET)-hours per week were then estimated by multiplying the MET score for each activity by the reported hours per week and summing across all activities as previously described.[9, 24] In the NHS, women responded to a validated food frequency questionnaire (FFQ) in 1984. Similar FFQs were subsequently administered every 4 years in the NHS beginning in 1986. In the NHSII, women completed an FFQ in 1991, with similar FFQs administered every 4 years afterwards. Vegetable servings per day were estimated from the FFQs.[15, 16] As mammograms were collected from only a subset of women

within the breast cancer case-control studies nested in the NHS/NHSII[25, 26], we derived predicted percent MD using data from the NHS and NHSII nested case-control studies of breast cancer as previously described. [Rice et al., *Cancer Causes and Control*, in press] Briefly, using questionnaire data and mammograms from 1436 cases and 2955 controls, we assessed the following variables as potential predictors: age, adolescent somatotype, BMI at age 18, current BMI, age at menarche, parity, age at first birth, height, family history of breast cancer, personal history of BBD (confirmed by biopsy and unconfirmed), alcohol use, HT use, and duration of menopause. We also included interaction terms for each of the candidate predictors and menopausal status as potential contributors to the model. In addition, menopausal status, nulliparity, and breast cancer case/control status were forced into the model. Percent MD was square-root transformed to improve normality and stepwise linear regression ($p < 0.15$ for forward and backward steps) was used to select the subset of candidate variables that were most predictive of square-root transformed percent MD. The following variables were selected for inclusion (in addition to menopausal status, nulliparity, and case/control status): age, current BMI, BMI at age 18, HT use, biopsy confirmed BBD, unconfirmed BBD, adolescent somatotype, parity, and age at first birth as well as the interaction term between menopausal status and age at first birth. The final percent MD prediction model explained 42% of the total variability in square-root transformed percent MD.

Breast cancer cases

Breast cancer cases were identified on the biennial questionnaires or through death records. These participants or their next of kin were asked for permission to obtain their medical records. We obtained pathology reports from 90% of cases and information on histologic tumor type, tumor size, and nodal involvement was reviewed and extracted by study investigators. As we confirm 99% of reported cases of breast cancer for whom we are able to obtain medical records, all diagnoses of invasive breast cancer confirmed by the participant or medical records were included as cases in our analysis. ER and PR status of the tumors was determined either by immunohistochemical staining of participant tumor tissue as described previously or, when tissue was unavailable, by medical record abstraction.[27]

Population for analysis

For all analyses, we initiated follow-up in 1986 in the NHS and 1989 in the NHSII. Exclusions at the start of follow-up are presented in Figure 1. Women were allowed to enter the analysis at the time they first completed a FFQ or physical activity assessment later in follow-up. Throughout follow-up, if a woman did not complete the most recent FFQ or questions on physical activity she did not contribute person-time until she answered a subsequent FFQ or set of physical activity items. If a woman stopped reporting her current weight on the biennial questionnaires, she was censored from the analysis at the time of last report. Further, women were censored if they newly met the exclusion criteria (e.g., reported onset of menopause due to reasons other than natural menopause or bilateral oophorectomy). At baseline, 133,640 women were included in the analysis.

Description of model

The Rosner-Colditz model for breast cancer incidence has been previously described in detail.[1, 2] Here, women were censored when they developed other types of cancer (except non-melanoma skin cancer), if they died, or at the end of follow-up (NHS=2010, NHSII=2009), whichever occurred first. We used PROC LIFEREG in SAS to fit an exponential survival model based on the following Rosner-Colditz risk model factors: 1) duration of premenopause, 2) duration of postmenopause by type of menopause (natural or surgical), 3) time between age at menarche and age at first birth, 4) birth index, 5) past or current HT use, 6) duration of HT use by type (E, E+P, other), 7) BMI during premenopause and during postmenopause, 8) height during premenopause and during postmenopause, 9) alcohol intake during premenopause, during postmenopause on and not on HT, 10) family history of breast cancer, and 11) BBD (biopsy-confirmed and unconfirmed) including interaction terms with a) age at menarche, b) duration of premenopause, and c) duration of postmenopause.[1, 2] In addition, we adjusted for cohort (NHSII vs NHS) to allow for differing baseline risks between the two cohorts. While prior derivations of the Rosner-Colditz model used PROC PHREG, PROC LIFEREG estimates an intercept allowing for the calculation of absolute risks. As a sensitivity analysis, we also fit the Rosner-Colditz model using PROC PHREG and beta estimates and standard errors were very similar to those estimated with PROC LIFEREG (data not shown).

Statistical analysis

We added each new potential variable for inclusion (i.e., adolescent somatotype, physical activity, vegetable intake, breastfeeding, and predicted percent MD) to each breast cancer incidence model (i.e., all invasive, ER+/PR+, ER+/PR-, ER-/PR-) individually. We then compared the age-adjusted area under the curve (AUC) for the models without and with the new variable to assess improvement in risk prediction.[28] In our primary analysis (all invasive breast cancer), variables that significantly improved risk prediction (change in AUC $p < 0.05$) were added to the final updated model. In ER/PR subtype analysis, due to limited power for some subtypes, new candidate predictors were maintained if the p-value for the effect estimate in the model was < 0.05 and improved the age-adjusted AUC by at least 0.001, regardless of the p-value for change in AUC. We compared the age-adjusted AUCs from the final updated models to the original model.

We used 10-fold cross-validation to assess overfitting of the model. First, the dataset was divided into 10 equal sized groups. Next, the original and updated models were fit to a population comprised of 9 of the groups. In the 10th group, we estimated risk scores for the models with and without the new variables by summing over the cross products of coefficients of the given risk model (β_i) from the population of 9 groups and values of the risk factors in the 10th group (X_i) ($score = \sum \beta_i X_i$) to assess change in AUC. This process was repeated 9 more times. We then averaged the change in AUC across the 10 analyses and estimated the average change in AUC using fixed-effects meta-analysis.[29]

We also computed age-specific deciles of the risk score for all invasive breast cancer for the model without (model 1) and with (model 2) the newly included variables. We then cross-classified risk deciles from model 1 with risk deciles from model 2 and estimated the

incidence rate in each strata. Specifically, let X_{ij} = number of breast cancer cases, N_{ij} = number of person-years and $\hat{P}_{ij} = X_{ij}/N_{ij}$ estimated incidence rate within the i th age – specific risk decile for model 1 and the j th age-specific risk decile for model 2 and let $\ln(P_{ij}) = \alpha_i + \beta(j-1)$. [30] $\exp(\beta)$ is an estimate of the percent increase in breast cancer incidence for an increase of one model 2 risk decile, holding the model 1 risk decile constant.

To evaluate calibration, we grouped women by decile of absolute risk and estimated the observed (O) and expected (E) counts of incident breast cancer cases in each decile (d). We then ran the following Poisson regression to assess goodness of fit:

$$\ln(\mu^{(d)}) = \alpha + \ln(E^{(d)}) \text{ where } \mu^{(d)} = E(O^{(d)})$$

where we tested the hypotheses $H_0: \alpha = 0$ vs. $H_1: \alpha \neq 0$. [31]

We also estimated the Hosmer-Lemeshow statistic:

$$X_{HL}^2 = \sum_{d=1}^{10} \frac{[O^{(d)} - E^{(d)}]^2}{E^{(d)}}$$

and compared X_{HL}^2 to a chi-square distribution with 8 d.f. All p-values are two-sided and were considered statistically significant if <0.05 . All analyses were conducted in SAS statistical software version 9.4 (Cary, North Carolina).

Results

Table 1 presents participant characteristics by cohort at the midpoint of follow-up (NHS=1996; NHSII=1997). Throughout follow-up we identified 5,246 cases of incident invasive breast cancer. Of these, 3,154 were classified as ER+/PR+, 606 were ER+/PR–, and 775 were ER–/PR–; the remaining 711 tumors were unclassifiable.

Both adolescent somatotype and predicted percent MD individually significantly improved the original Rosner-Colditz model for all invasive breast cancer (Supplementary table 1) whereas vegetable intake, breastfeeding, and physical activity did not. The age-adjusted AUC for the model including both adolescent somatotype and predicted percent MD was 0.636 compared to 0.616 for the model without these variables (change in age-adjusted AUC=0.020, $p<0.0001$) (Table 2). While the AUC for the updated model was relatively consistent across strata of age, it was somewhat lower among women aged ≥ 70 (AUC_{ages 70–74} = 0.609 and AUC_{ages ≥ 75} = 0.610). The relative risk (RR) of invasive breast cancer for a 4-unit increase in adolescent somatotype was 0.62 (95%CI: 0.56, 0.70) and the RR for a 20-unit increase in predicted percent MD was 1.32 (95%CI: 1.28, 1.36). The inclusion of these two risk factors did not substantially change the estimates for other risk factors in the model with the exception of premenopausal BMI. Premenopausal BMI was inversely associated with breast cancer risk in the original model (RR_{per 296 unit increase} =

0.82; 95%CI: 0.76, 0.88), but positively associated with risk in the updated model ($RR_{\text{per 296 unit increase}} = 1.17$; 95%CI: 1.08, 1.27) that also included adolescent somatotype and predicted percent MD. Results from the 10-fold cross-validation were consistent with our initial observation, with a fixed effects change in age-adjusted AUC of 0.020 ($p < 0.0001$). Cross-classifying the basic Rosner-Colditz model with the updated model including adolescent somatotype and predicted percent MD suggested that there were significant differences in estimated incidence of invasive breast cancer (Table 3). The overall slope was 0.168 ($p < 0.001$) indicating a significant 16.8% increase in invasive breast cancer incidence for an increase of one age-specific risk decile for the updated model, holding the age-specific basic model risk decile constant. Therefore, adolescent somatotype and predicted percent MD significantly increased the predictive power of the Rosner-Colditz model. Calibration by the Hosmer-Lemeshow test ($p = 0.68$) and Poisson regression ($p = 0.99$) indicated adequate fit (Table 4).

Similar to the overall model, adolescent somatotype and predicted percent MD also significantly improved the ER+/PR+ model (age-adjusted $AUC_{\text{basic model}} = 0.629$ vs age-adjusted $AUC_{\text{updated model}} = 0.651$; $p < 0.001$) as well as the ER+/PR- model (age-adjusted $AUC_{\text{basic model}} = 0.627$ vs age-adjusted $AUC_{\text{updated model}} = 0.639$; $p = 0.007$) (Table 5). Adolescent somatotype, predicted percent MD, breastfeeding, and vegetable intake individually significantly improved the ER-/PR- model (Supplementary table 1). The age-adjusted AUC for the model including these all four of these risk factors was 0.630 compared to 0.598 for the model without these variables (change in age-adjusted $AUC = 0.031$, $p < 0.0001$) (Table 5). In the ER-/PR- model, the RR for a 4-unit increase in adolescent somatotype was 0.46 (95%CI: 0.34, 0.62). In addition, the RR for 5 vegetable servings/day was 0.83 (95%CI: 0.70, 0.99), while the RR for 12 months of breastfeeding was 0.88 (95%CI: 0.77, 1.01) in the ER-/PR- model. Physical activity did not improve risk classification beyond the existing factors in the model (Supplemental table 1).

Discussion

We observed that adolescent somatotype and predicted percent MD improved the discriminatory ability of the Rosner-Colditz model for all invasive breast cancer as well as for models defined by ER/PR status. In addition, vegetable intake and breastfeeding duration also significantly contributed to improving the model for ER-/PR- breast cancer, but not ER+ models, consistent with prior research [17,18,19]. The age-adjusted AUC for the final model for all invasive breast cancer was 0.636.

In prior analyses in the NHS/NHSII, we observed an inverse association between adolescent somatotype and breast cancer risk, which was stronger for ER- tumors than ER+ tumors and persisted after adjustment for adult BMI. [7, 8, 32] Consistent with these earlier observations, the inverse association between adolescent somatotype and risk was stronger in the Rosner-Colditz model for ER-/PR- breast cancer compared to the association in either the ER+/PR+ model or the ER+/PR- model. While this inverse association has been observed in other populations,[3–6] the mechanism by which increased body size in adolescence lowers risk is not well understood. Hypothesized mechanisms include earlier

differentiation of breast tissue,[33] altered hormone levels (e.g., lower IGF-1), [34] and lower percent MD in adulthood, among others.[35]

Higher percent MD, representing a greater proportion of epithelial and stromal tissue on a mammogram, is one of the strongest risk factors, though only a few studies have evaluated the role of MD in risk prediction models. [36] The Breast Cancer Surveillance Consortium (BCSC) risk calculator currently includes Breast Imaging-Reporting and Data System (BI-RADS) breast density, a four category radiologist assessment.[37, 38] While BI-RADS is predictive of subsequent breast cancer risk, variability in BI-RADS is limited (e.g., over 80% of women in the BCSC were BI-RADS 2 or 3) compared to continuous measures, such as percent MD. Recent work in both in the International Breast Cancer Intervention Study I as well as the Predicting Risk of Cancer at Screening study observed that including percent MD (adjusted for age and BMI) improved risk prediction using the Tyrer-Cuzick as well as the Gail model.[21, 22] Our observation that predicted percent MD improved discrimination for the Rosner-Colditz model, which is comprised of a greater number of breast cancer risk factors than many other breast cancer prediction models, is consistent with these prior studies and supports the inclusion of percent MD in future prediction models.

The original Rosner-Colditz model was developed to predict all invasive breast cancer; however since the majority of invasive disease is ER+, most of the risk factors included in the original model were more strongly associated with ER+ than ER- disease. As a result, the AUC for the original Rosner-Colditz model was lower for ER-/PR- breast cancer (age-adjusted AUC=0.598) compared to ER+ tumors (age-adjusted AUC=0.629 and 0.627, for ER+/PR+ and ER+/PR- respectively). Thus, the impact of including risk factors specific to (vegetable intake, breastfeeding) or stronger for (adolescent somatotype) ER-/PR- disease resulted in a larger increase in the AUC in the updated model for this subtype (change in age-adjusted AUC=0.031) compared to ER+ breast cancer (change in age-adjusted AUC=0.020 and 0.012, for ER+/PR+ and ER+/PR- respectively). Despite this, the overall predictive ability of the revised Rosner-Colditz model is still better for ER+/PR+ versus ER-/PR- tumors. This highlights the need to identify and consider subtype specific risk factors in prediction models, particularly for more aggressive subtypes.

We previously observed that physical activity was inversely associated with risk of breast cancer; however adding physical activity to the original Rosner-Colditz model did not significantly improve the any of the breast cancer models.[9] While others have also observed an inverse association with risk, physical activity may not contribute to risk prediction above and beyond other lifestyle factors, such as BMI across the lifecourse, already included in the Rosner-Colditz model.[10–12]

There are several strengths to our study, including the large sample size, prospectively collected data, repeated measurements of risk factors allowing for updating over time, and centralized pathology review and/or medical record abstraction for classification of tumors by ER/PR status. A limitation of the analysis is that some women were excluded from the study population due to missing information on the selected risk factors. Further, the NHS/NHSII are predominantly White; future research is necessary to determine the clinical utility of the updated models in diverse populations. Non-differential misclassification of the

included risk factors, particularly predicted percent MD, may have led to an underestimation of some associations; additional evaluation using measured percent MD may show further improvement in predictive capacity of the model. Lastly, due to limited data in the majority of the cohort, we were unable to assess the contributions to risk prediction of other breast cancer risk factors such as breast cancer SNPs[39] or circulating hormones.[40]

Inclusion of adolescent somatotype and predicted percent MD improved the discrimination of the Rosner-Colditz all invasive breast cancer risk model as well as models defined by ER/PR status. Further, vegetable intake and duration of breastfeeding contributed to the ER–/PR– model. Given that the discriminatory ability is still somewhat limited, adding newly identified risk factors as well as established biomarkers associated with breast cancer risk is critical to enhancing clinical impact. Importantly, continued focus on subtype specific models is crucial to identify women at high risk of ER+ disease who could benefit from chemoprevention for this subtype (e.g., SERMs) as well as those at high risk of ER– disease who may benefit from increased screening. Future work is necessary to validate the updated model in independent and diverse populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding/Support: This study was supported by research grants from the National Cancer Institute, National Institutes of Health, UM1 CA186107, P01 CA87969, UM1 CA176726, R01CA175080, and T32 CA09001. We would like to thank the participants of the Nurses' Health Study and Nurses' Health Study II for their continuing contributions. We thank 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.

Abbreviations

NHS	Nurses' Health Study
BMI	body mass index
BBD	benign breast disease
HT	hormone therapy
MD	mammographic density
ER	estrogen receptor
PR	progesterone receptor

References

1. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst.* 1996; 88(6):359–64. [PubMed: 8609645]

2. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am J Epidemiol.* 2000; 152(10):950–64. [PubMed: 11092437]
3. Li J, et al. Effects of childhood body size on breast cancer tumour characteristics. *Breast Cancer Res.* 2010; 12(2):R23. [PubMed: 20398298]
4. Ahlgren M, et al. Growth patterns and the risk of breast cancer in women. *N Engl J Med.* 2004; 351(16):1619–26. [PubMed: 15483280]
5. Bardia A, et al. Relative weight at age 12 and risk of postmenopausal breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2008; 17(2):374–8. [PubMed: 18250344]
6. Hilakivi-Clarke L, et al. Tallness and overweight during childhood have opposing effects on breast cancer risk. *Br J Cancer.* 2001; 85(11):1680–4. [PubMed: 11742488]
7. Baer HJ, et al. Body fatness during childhood and adolescence and incidence of breast cancer in premenopausal women: a prospective cohort study. *Breast Cancer Res.* 2005; 7(3):R314–25. [PubMed: 15987426]
8. Baer HJ, et al. Body fatness at young ages and risk of breast cancer throughout life. *Am J Epidemiol.* 2010; 171(11):1183–94. [PubMed: 20460303]
9. Eliassen AH, et al. Physical activity and risk of breast cancer among postmenopausal women. *Arch Intern Med.* 2010; 170(19):1758–64. [PubMed: 20975025]
10. Peters TM, et al. Physical activity and postmenopausal breast cancer risk in the NIH-AARP diet and health study. *Cancer Epidemiol Biomarkers Prev.* 2009; 18(1):289–96. [PubMed: 19124511]
11. Dallal CM, et al. Long-term recreational physical activity and risk of invasive and in situ breast cancer: the California teachers study. *Arch Intern Med.* 2007; 167(4):408–15. [PubMed: 17325304]
12. Steindorf K, et al. Physical activity and risk of breast cancer overall and by hormone receptor status: the European prospective investigation into cancer and nutrition. *Int J Cancer.* 2013; 132(7):1667–78. [PubMed: 22903273]
13. Wu Y, Zhang D, Kang S. Physical activity and risk of breast cancer: a meta-analysis of prospective studies. *Breast Cancer Res Treat.* 2013; 137(3):869–82. [PubMed: 23274845]
14. Neilson HK, et al. Moderate-vigorous recreational physical activity and breast cancer risk, stratified by menopause status: a systematic review and meta-analysis. *Menopause.* 2017; 24(3):322–344. [PubMed: 27779567]
15. Fung TT, et al. Dietary patterns and the risk of postmenopausal breast cancer. *Int J Cancer.* 2005; 116(1):116–21. [PubMed: 15756679]
16. Fung TT, et al. Intake of specific fruits and vegetables in relation to risk of estrogen receptor-negative breast cancer among postmenopausal women. *Breast Cancer Res Treat.* 2013; 138(3):925–30. [PubMed: 23532538]
17. Jung S, et al. Fruit and vegetable intake and risk of breast cancer by hormone receptor status. *J Natl Cancer Inst.* 2013; 105(3):219–36. [PubMed: 23349252]
18. Sisti JS, et al. Reproductive factors, tumor estrogen receptor status and contralateral breast cancer risk: results from the WECARE study. *Springerplus.* 2015; 4:825. [PubMed: 26751177]
19. Sisti JS, et al. Reproductive risk factors in relation to molecular subtypes of breast cancer: Results from the Nurses' Health Studies. *Int J Cancer.* 2015; 18(10):29968.
20. Boyd NF, et al. Mammographic density. *Breast Cancer Res.* 2009; 11(Suppl 3):S4. [PubMed: 20030879]
21. Warwick J, et al. Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: findings from the placebo arm of the International Breast Cancer Intervention Study I. *Breast Cancer Res.* 2014; 16(5):451. [PubMed: 25292294]
22. Brentnall AR, 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. [PubMed: 26627479]
23. Harris HR, et al. Body size across the life course, mammographic density, and risk of breast cancer. *Am J Epidemiol.* 2011; 174(8):909–18. [PubMed: 21911827]

24. Colditz GA, et al. Physical activity and risk of breast cancer in premenopausal women. *Br J Cancer*. 2003; 89(5):847–51. [PubMed: 12942116]
25. Rice MS, et al. Mammographic density and breast cancer risk: a mediation analysis. *Breast Cancer Res*. 2016; 18(1):94. [PubMed: 27654859]
26. Yaghjian L, et al. Postmenopausal mammographic breast density and subsequent breast cancer risk according to selected tissue markers. *Br J Cancer*. 2015; 113(7):1104–13. [PubMed: 26335607]
27. Tamimi RM, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. *Breast Cancer Res Treat*. 2012; 131(1):159–67. [PubMed: 21830014]
28. Rosner B, Glynn RJ. Power and sample size estimation for the Wilcoxon rank sum test with application to comparisons of C statistics from alternative prediction models. *Biometrics*. 2009; 65(1):188–97. [PubMed: 18510654]
29. DerSimonian R. Meta-analysis in the design and monitoring of clinical trials. *Stat Med*. 1996; 15(12):1237–48. discussion 1249–52. [PubMed: 8817798]
30. Rosner B, et al. Risk prediction models with incomplete data with application to prediction of estrogen receptor-positive breast cancer: prospective data from the Nurses' Health Study. *Breast Cancer Res*. 2008; 10(4):R55. [PubMed: 18598349]
31. Rosner BA, et al. Validation of Rosner-Colditz breast cancer incidence model using an independent data set, the California Teachers Study. *Breast Cancer Res Treat*. 2013; 142(1):187–202. [PubMed: 24158759]
32. Berkey CS, et al. Adolescence and breast carcinoma risk. *Cancer*. 1999; 85(11):2400–9. [PubMed: 10357411]
33. Baer HJ, et al. Adiposity and sex hormones in girls. *Cancer Epidemiol Biomarkers Prev*. 2007; 16(9):1880–8. [PubMed: 17855709]
34. Poole EM, et al. Body size in early life and adult levels of insulin-like growth factor 1 and insulin-like growth factor binding protein 3. *Am J Epidemiol*. 2011; 174(6):642–51. [PubMed: 21828371]
35. Samimi G, et al. Measures of energy balance and mammographic density in the Nurses' Health Study. *Breast Cancer Res Treat*. 2008; 109(1):113–22. [PubMed: 17592770]
36. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2006; 15(6):1159–69. [PubMed: 16775176]
37. Tice JA, et al. 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. [PubMed: 18316752]
38. Vachon CM, et al. The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst*. 2015; 107(5)
39. Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013; 45(4):353–61. 361e1–2. [PubMed: 23535729]
40. Tworoger SS, et al. Inclusion of endogenous hormone levels in risk prediction models of postmenopausal breast cancer. *J Clin Oncol*. 2014; 32(28):3111–7. [PubMed: 25135988]

Figure 1.
Study population exclusions at start of follow-up (NHS=1986; NHSII=1989)

Table 1

Characteristics of the study population at the midpoint of follow-up, in 1996 (NHS) and 1997 (NHSII)

	NHS (N=40,488)	NHSII (N=59,338)
Mean (SD)		
Age (years)	62.3(7.2)	42.4(4.4)
Age at menarche (years)	12.6(1.4)	12.4(1.4)
Adolescent somatotype	2.7(1.2)	2.9(1.1)
BMI (kg/m ²)	26.4(5.2)	26.1(6.0)
Height (inches)	64.5(2.4)	64.9(2.6)
Parity *	3.1(1.5)	2.2(0.9)
Age at first birth (years) *	25.2(3.3)	26.1(4.4)
Birth Index *	66.7(34.5)	31.6(18.3)
Breastfeeding (months) *	6.4(10.6)	14.5(14.3)
Alcohol (grams/day)	4.8(8.0)	3.3(6.0)
Vegetable intake (servings/day)	3.7(2.1)	3.8(2.3)
Physical activity (MET-hrs/week)	17.8(20.3)	18.1(20.5)
Predicted percent MD	23.6(16.4)	40.5(21.1)
Years E only HT use ^	2.6(5.5)	3.1(4.2)
Years E+P HT use ^	1.6(2.8)	0.9(2.2)
Years other HT use ^	0.7(1.9)	0.1(0.7)
Percent		
Nulliparous	5.9	22.5
Family history of breast cancer	14.2	9.7
Biopsy confirmed BBD	20.6	16.2
Unconfirmed BBD	26.8	33.7
Postmenopausal	92.4	10.4
Type of Menopause ^		
Natural	79.6	39.3
Bilateral oophorectomy	20.4	60.7
Hormone therapy use ^		
Past	20.8	8.2
Current	45.3	78.2

BMI=body mass index, MET=metabolic equivalent task, E = estrogen, E+P = estrogen and progesterone, HT = hormone therapy, BBD=benign breast disease, MD=mammographic density

* Parous only

^ Postmenopausal only

Fitted invasive breast cancer Rosner-Colditz incidence model, NHS (1986–2010) and NHSII (1989–2009)

Table 2

	Increment	Original model			Updated model		
		HR (95%CI)	parameter	SE	HR (95%CI)	parameter	SE
Intercept			-9.2797			-9.7161	
Cohort (NHSII vs NHS)	1	0.88 (0.80, 0.96)	-0.1307	0.0446	0.86 (0.79, 0.94)	-0.1469	0.0449
Duration of premenopause (yrs)	1	1.09 (1.08, 1.10)	0.0846	0.0041	1.10 (1.09, 1.10)	0.0915	0.0042
Duration of menopause (yrs)							
Natural	1	1.02 (1.01, 1.02)	0.0174	0.0028	1.03 (1.02, 1.03)	0.0254	0.0028
Bilateral oophorectomy	1	1.02 (1.01, 1.02)	0.0166	0.0039	1.03 (1.02, 1.03)	0.0252	0.0039
Pregnancy history							
Age at 1st birth – age at menarche (yrs)	22	1.16 (1.04, 1.29)	0.0066	0.0025	1.15 (1.03, 1.28)	0.0063	0.0024
Birth index	102	0.71 (0.65, 0.78)	-0.0034	0.0005	0.76 (0.69, 0.84)	-0.0027	0.0005
Biopsy-confirmed BBD							
BBD (yes v no)	1	1.64 (0.82, 3.31)	0.4972	0.3576	1.68 (0.83, 3.39)	0.5166	0.3591
BBD*age at menarche	1	1.04 (1.00, 1.08)	0.0401	0.0199	1.03 (0.99, 1.07)	0.0286	0.0200
BBD*(duration of premenopause)	1	0.99 (0.98, 1.00)	-0.0125	0.0061	0.99 (0.98, 1.00)	-0.0121	0.0061
BBD*(duration of menopause)	1	0.99 (0.98, 1.00)	-0.0107	0.0033	0.99 (0.98, 1.00)	-0.0090	0.0033
Unconfirmed BBD							
BBD (yes v no)	1	1.50 (0.80, 2.79)	0.4046	0.3179	1.69 (0.90, 3.17)	0.5266	0.3197
BBD*age at menarche	1	1.02 (0.99, 1.06)	0.0215	0.0182	1.01 (0.97, 1.04)	0.0056	0.0184

All invasive breast cancer Cases/person-years=5,246/2,063,532						
	Original model			Updated model		
	Increment	HR (95%CI)	parameter	SE	HR (95%CI)	parameter
BBD*(duration of premenopause)	1	0.99 (0.98, 1.00)	-0.0106	0.0055	0.99 (0.98, 1.00)	-0.0103
BBD*(duration of menopause)	1	0.99 (0.99, 1.00)	-0.0057	0.0033	1.00 (0.99, 1.00)	-0.0039
Postmenopausal HT use						
Duration oral estrogen	10	1.16 (1.06, 1.27)	0.0150	0.0047	1.14 (1.04, 1.25)	0.0128
Duration oral estrogen and progesterone	10	1.48 (1.33, 1.65)	0.0393	0.0053	1.47 (1.32, 1.63)	0.0385
Duration other HT use	10	1.31 (1.10, 1.55)	0.0267	0.0089	1.27 (1.06, 1.51)	0.0237
Current HT use	1	1.26 (1.16, 1.38)	0.2329	0.0445	1.26 (1.15, 1.37)	0.2272
Past HT use	1	0.86 (0.79, 0.94)	-0.1521	0.0455	0.88 (0.80, 0.96)	-0.1333
BMI						
(Average BMI during premenopause-21.8)*(duration of premenopause)	296	0.82 (0.76, 0.88)	-0.0007	0.0001	1.17 (1.08, 1.27)	0.0005
(Average BMI after menopause-24.4)*(duration of menopause)	160	1.52 (1.38, 1.67)	0.0026	0.0003	1.51 (1.37, 1.66)	0.0026
Height						
(Height-64.5)*duration of premenopause	222	1.17 (1.09, 1.25)	0.0007	0.0002	1.18 (1.10, 1.26)	0.0007
(Height-64.4)*duration of menopause	120	1.00 (0.86, 1.17)	0.0000	0.0007	1.02 (0.87, 1.19)	0.0002
Alcohol						
Cumulative grams before menopause	352	1.19 (1.10, 1.27)	0.0005	0.0001	1.19 (1.11, 1.28)	0.0005
Cumulative grams after menopause						
with use of hormones	220	0.91 (0.83, 0.99)	-0.0004	0.0002	0.92 (0.84, 1.00)	-0.0004
without use of hormones	220	1.13 (1.05, 1.22)	0.0006	0.0002	1.15 (1.06, 1.23)	0.0006

All invasive breast cancer Cases/person-years=5,246/2,063,532							
	Original model			Updated model			SE
	Increment	HR (95%CI)	parameter	SE	HR (95%CI)	parameter	
Family history of breast cancer (yes v no)	1	1.53 (1.43,1.64)	0.4258	0.0356	1.54 (1.43,1.65)	0.4292	0.0356
Adolescent somatotype	4				0.62 (0.56,0.70)	-1.180	0.0143
Predicted percent MD	20				1.32 (1.28,1.36)	0.0139	0.0008
<i>Age group</i>	<i>Cases/PY</i>		<i>AUC (SE)</i>			<i>AUC (SE)</i>	<i>Difference in AUCs</i>
<40	238/446,784		0.6577 (0.0175)			0.6708 (0.0173)	0.0131 (0.0108)
40-44	507/336,756		0.6229 (0.0123)			0.6345 (0.0122)	0.0116 (0.0082)
45-49	718/305,310		0.5974 (0.0105)			0.6333 (0.0103)	0.0358 (0.0065)
50-54	654/234,079		0.6121 (0.0109)			0.6377 (0.0108)	0.0256 (0.0062)
55-59	718/190,276		0.6254 (0.0104)			0.6451 (0.0102)	0.0197 (0.0052)
60-64	747/183,248		0.6137 (0.0102)			0.6305 (0.0101)	0.0168 (0.0049)
65-69	710/158,380		0.6385 (0.0103)			0.6573 (0.0102)	0.0188 (0.0048)
70-74	527/112,811		0.5914 (0.0123)			0.6090 (0.0122)	0.0175 (0.0056)
75+	427/95,888		0.5918 (0.0137)			0.6103 (0.0136)	0.0185 (0.0060)
Overall	5,246/2,063,532		0.7015 (0.0036)			0.7139 (0.0035)	0.0124 (0.0013) <0.0001
Overall: age-adjusted	5,246/2,063,532		0.6155 (0.0039)			0.6362 (0.0038)	0.0201 (0.0020) p<0.0001

YRS=years, BMI = body mass index, HT = hormone therapy, BBD = benign breast disease, MD=mammographic density AUC = area under the curve, PY = person-years, SE = standard error

Table 3

Cross-classification of original model risk decile by updated model risk decile

Updated model risk decile													
	1	2	3	4	5	6	7	8	9	10	Slope	SE	P-value
Basic model risk decile	Cases/Person-years (Rate)	Cases/Person-years (Rate)	Cases/Person-years (Rate)	Cases/Person-years (Rate)	Cases/Person-years (Rate)	Cases/Person-years (Rate)	Cases/Person-years (Rate)	Cases/Person-years (Rate)	Cases/Person-years (Rate)	Cases/Person-years (Rate)			
1	132/131,142 (101)	56/41,887 (134)	32/16,957 (189)	13/8,191 (159)	4/4,275 (94)	1/2,335 (43)	2/1,070 (187)	2/394 (508)	0/96 (0)	0/2 (0)	0.156	0.051	0.002
2	45/49,134 (92)	101/66,263 (152)	67/40,202 (167)	45/23,163 (194)	21/13,109 (160)	11/7,555 (146)	13/4,244 (306)	2/2,031 (98)	0/662 (0)	0/30 (0)	0.117	0.036	0.001
3	19/17,747 (107)	74/49,285 (150)	94/49,565 (190)	64/36,237 (177)	62/23,507 (264)	30/14,381 (209)	26/8,625 (301)	9/4,845 (186)	1/1,955 (51)	1/209 (478)	0.109	0.029	<0.001
4	3/5,971 (50)	34/28,180 (121)	62/43,717 (142)	104/42,689 (244)	78/33,913 (230)	48/23,564 (204)	45/14,895 (302)	20/8,505 (235)	10/4,252 (235)	4/670 (597)	0.125	0.027	<0.001
5	0/1,792 (0)	9/13,554 (66)	44/29,606 (149)	69/40,011 (172)	85/40,512 (210)	92/33,075 (278)	56/23,531 (238)	51/14,622 (349)	16/7,845 (204)	7/1,802 (388)	0.145	0.027	<0.001
6	1/465 (215)	2/5,318 (38)	26/16,779 (155)	50/29,643 (169)	87/38,693 (227)	110/40,040 (275)	68/33,415 (204)	78/24,100 (324)	41/13,614 (301)	20/4,286 (467)	0.127	0.025	<0.001
7	0/80 (0)	1/1503 (67)	6/7128 (84)	34/17,507 (194)	68/29,918 (227)	92/39,771 (231)	143/42,263 (338)	128/35,797 (358)	109/23,654 (461)	31/8,736 (355)	0.168	0.025	<0.001
8	0/18 (0)	1/329 (304)	2/2,130 (94)	6/7,343 (82)	26/16,756 (155)	63/29,693 (212)	128/42,938 (298)	178/47,921 (371)	166/40,299 (412)	86/18,928 (454)	0.204	0.028	<0.001
9	0/0 (-)	0/34 (0)	0/262 (0)	1/1489 (67)	12/5224 (230)	23/14,094 (163)	83/28,916 (287)	146/48,947 (298)	256/62,675 (408)	217/44,711 (485)	0.210	0.031	<0.001
10	0/0 (-)	0/0 (-)	0/12 (0)	0/80 (0)	0/442 (0)	1/1,850 (54)	10/6,457 (155)	48/19,194 (250)	198/51,339 (386)	736/126,977 (580)	0.426	0.051	<0.001
Overall											0.168	0.005	<0.001

SE=standard error

Table 4

Calibration of the updated Rosner-Colditz Model

	1	2	3	4	5	6	7	8	9	10
Updated model risk decile										
Observed number of cases	200	278	333	386	443	471	574	662	797	1102
Expected number of cases	196	273	325	374	425	483	552	643	784	1191
Hosmer-Lemeshow χ^2	6.62									
p-value	0.58									
Poisson regression α	0.0000									
SE (α)	0.0138									
p-value	0.99									

SE=standard error

Fitted incidence models for breast cancer by ER/PR status, NHS (1986–2010) and NHSII (1989–2009)

Table 5

	Increment	ER+/PR+		ER+/PR–		ER–/PR–	
		Original model HR (95%CI)	Updated model HR (95%CI)	Original Model HR (95%CI)	Updated model HR (95%CI)	Original model HR (95%CI)	Updated model HR (95%CI)
Cohort (NHSII vs NHS)	1	0.95 (0.84, 1.06)	0.93 (0.83, 1.04)	0.60 (0.46, 0.78)	0.58 (0.44, 0.77)	0.83 (0.66, 1.04)	0.88 (0.70, 1.10)
Duration of premenopause (yrs)	1	1.10 (1.09, 1.11)	1.11 (1.10, 1.12)	1.09 (1.06, 1.12)	1.10 (1.07, 1.13)	1.06 (1.04, 1.08)	1.06 (1.04, 1.08)
Duration of menopause (yrs)							
Natural	1	1.02 (1.01, 1.03)	1.03 (1.02, 1.04)	1.02 (1.00, 1.04)	1.03 (1.01, 1.04)	1.01 (0.99, 1.02)	1.01 (1.00, 1.03)
Bilateral oophorectomy	1	1.02 (1.01, 1.03)	1.03 (1.02, 1.04)	1.01 (0.99, 1.03)	1.02 (1.00, 1.04)	1.01 (0.99, 1.03)	1.02 (1.00, 1.04)
Pregnancy history							
Age at 1st birth – age at menarche (yrs)	22	1.06 (0.93, 1.22)	1.05 (0.92, 1.21)	1.70 (1.22, 2.37)	1.67 (1.21, 2.32)	1.43 (1.07, 1.90)	1.53 (1.14, 2.05)
Birth index	102	0.67 (0.59, 0.75)	0.72 (0.64, 0.81)	0.70 (0.53, 0.91)	0.74 (0.57, 0.97)	1.02 (0.80, 1.30)	1.14 (0.88, 1.48)
Biopsy-confirmed BBD							
BBD (yes v no)	1	1.52 (0.61, 3.78)	1.51 (0.61, 3.78)	0.26 (0.03, 2.47)	0.26 (0.03, 2.48)	2.83 (0.47, 17.06)	3.38 (0.56, 20.54)
BBD*age at menarche	1	1.04 (0.99, 1.10)	1.03 (0.98, 1.09)	1.14 (1.02, 1.28)	1.13 (1.01, 1.27)	0.98 (0.88, 1.09)	0.96 (0.86, 1.06)
BBD*(duration of premenopause)	1	0.99 (0.97, 1.01)	0.99 (0.98, 1.01)	1.00 (0.97, 1.04)	1.01 (0.97, 1.05)	0.99 (0.96, 1.02)	0.99 (0.96, 1.02)
BBD*(duration of menopause)	1	0.98 (0.98, 0.99)	0.99 (0.98, 0.99)	0.99 (0.97, 1.01)	0.99 (0.97, 1.01)	1.01 (0.99, 1.02)	1.01 (0.99, 1.03)
Unconfirmed BBD							
BBD (yes v no)	1	0.97 (0.43, 2.19)	1.08 (0.48, 2.44)	1.76 (0.24, 12.90)	1.93 (0.26, 14.25)	3.43 (0.72, 16.25)	4.46 (0.93, 21.39)
BBD*age at menarche	1	1.06 (1.01, 1.11)	1.04 (0.99, 1.09)	0.96 (0.86, 1.07)	0.95 (0.85, 1.06)	0.96 (0.87, 1.05)	0.93 (0.85, 1.02)
BBD*(duration of premenopause)	1	0.99 (0.98, 1.00)	0.99 (0.98, 1.00)	1.00 (0.97, 1.04)	1.00 (0.97, 1.04)	0.99 (0.96, 1.02)	0.99 (0.96, 1.02)
BBD*(duration of menopause)	1	0.99 (0.98, 1.00)	0.99 (0.99, 1.00)	1.00 (0.99, 1.02)	1.00 (0.99, 1.02)	0.99 (0.97, 1.01)	0.99 (0.97, 1.01)

		ER+/PR+ Cases/PY =3,154/2,063,532		ER+/PR- Cases/PY =606/2,063,532		ER-/PR- Cases/PY =775/2,063,532	
	Increment	Original model HR (95%CI)	Updated model HR (95%CI)	Original Model HR (95%CI)	Updated model HR (95%CI)	Original model HR (95%CI)	Updated model HR (95%CI)
Postmenopausal HT use							
Duration oral estrogen	10	1.16 (1.03, 1.31)	1.13 (1.00,1.27)	1.25 (0.98,1.61)	1.24 (0.97,1.59)	1.04 (0.82,1.32)	1.02 (0.80,1.30)
Duration oral estrogen and progesterone	10	1.70 (1.49,1.93)	1.68 (1.48,1.91)	1.15 (0.85,1.54)	1.15 (0.85,1.54)	1.22 (0.91,1.63)	1.23 (0.92,1.64)
Duration other HT use	10	1.40 (1.12,1.74)	1.35 (1.08,1.69)	1.00 (0.59,1.67)	0.98 (0.58,1.64)	1.57 (1.04,2.39)	1.57 (1.03,2.38)
Current HT use	1	1.26 (1.13, 1.41)	1.25 (1.12,1.40)	1.35 (1.05,1.74)	1.33 (1.03,1.71)	1.30 (1.03,1.63)	1.30 (1.03,1.63)
Past HT use	1	0.81 (0.72, 0.91)	0.83 (0.74,0.93)	1.13 (0.89,1.43)	1.14 (0.89,1.45)	1.00 (0.79,1.26)	1.02 (0.80,1.29)
BMI							
(Average BMI during premenopause-21.8)*(duration of premenopause)	296	0.83 (0.76, 0.90)	1.19 (1.08,1.32)	0.71 (0.57,0.88)	0.93 (0.73,1.19)	0.83 (0.69,1.00)	1.29 (1.04,1.59)
(Average BMI after menopause-24.4)*(duration of menopause)	160	1.66 (1.48, 1.87)	1.65 (1.47,1.85)	1.42 (1.08,1.87)	1.44 (1.09,1.88)	1.00 (0.74,1.35)	0.98 (0.73,1.31)
Height							
(Height-64.5)*duration of premenopause	222	1.24 (1.13, 1.35)	1.25 (1.14,1.36)	0.95 (0.78,1.16)	0.96 (0.79,1.17)	1.22 (1.02,1.46)	1.25 (1.05,1.49)
(Height-64.4)*duration of menopause	120	1.01 (0.83, 1.23)	1.03 (0.85,1.25)	1.24 (0.81,1.90)	1.25 (0.82,1.91)	0.94 (0.61,1.44)	0.94 (0.62,1.44)
Alcohol							
Cumulative grams before menopause	352	1.17 (1.07, 1.29)	1.18 (1.07,1.29)	1.45 (1.19,1.77)	1.45 (1.19,1.77)	1.02 (0.83,1.26)	1.04 (0.84,1.27)
Cumulative grams after menopause							
with use of hormones	220	0.93 (0.83, 1.04)	0.94 (0.84,1.05)	0.92 (0.73,1.16)	0.93 (0.74,1.17)	0.83 (0.64,1.07)	0.85 (0.65,1.10)
without use of hormones	220	1.16 (1.05, 1.28)	1.17 (1.07,1.29)	1.08 (0.89,1.32)	1.09 (0.89,1.33)	1.07 (0.86,1.34)	1.08 (0.87,1.35)
Family history (yes v no)	1	1.56 (1.43, 1.71)	1.57 (1.44,1.72)	1.53 (1.25,1.88)	1.54 (1.26,1.88)	1.45 (1.20,1.74)	1.45 (1.21,1.75)
Somatotype (Ages 10+20)	4		0.63 (0.55,0.73)		0.73 (0.53,1.01)		0.46 (0.34,0.62)

	Increment	ER+/PR+ Cases/PY=3,154/2,063,532		ER+/PR- Cases/PY=606/2,063,532		ER-/PR- Cases/PY=775/2,063,532	
		Original model HR (95%CI)	Updated model HR (95%CI)	Original Model HR (95%CI)	Updated model HR (95%CI)	Original model HR (95%CI)	Updated model HR (95%CI)
Predicted percent MD	20		1.35 (1.29,1.40)		1.27 (1.15,1.40)		1.28 (1.18,1.39)
Vegetable Intake	5						0.83 (0.70,0.99)
Breastfeeding	12						0.88 (0.77,1.01)
AUC (SE)		0.7144 (0.0045)	0.7273 (0.0044)	0.7510 (0.0098)	0.7549 (0.0097)	0.6684 (0.0096)	0.6879 (0.0094)
Difference in AUCs		0.0129 (0.0016)	p<0.0001	0.0039 (0.0028)	p=0.15	0.0196 (0.0050)	<0.0001
Age-adjusted AUC (SE)		0.6294 (0.0049)	0.6512 (0.0048)	0.6270 (0.0112)	0.6393 (0.0111)	0.5980 (0.0101)	0.6301 (0.0099)
Difference in AUCs		0.0204 (0.0024)	p<0.0001	0.0115 (0.0043)	p=0.007	0.0311 (0.0069)	<0.0001

BMI = body mass index, HT = hormone therapy, BBD = benign breast disease, MD=mammographic density AUC = area under the curve, PY = person-years, SE = standard error