Article

Comparative validation of breast cancer risk prediction models and projections for future risk stratification

Parichoy Pal Choudhury,[1]* Amber N. Wilcox,[2]* Mark N. Brook,[3] Yan Zhang,[1] Thomas Ahearn,[2] Nick Orr,[4] Penny Coulson,[3] Minouk J. Schoemaker,[3] Michael E. Jones,[3] Mitchell H. Gail,[2] Anthony J. Swerdlow,[3,5] Nilanjan Chatterjee,[1,6]** Montserrat Garcia-Closas[2]**

Affiliations:

[1]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, USA

[2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, USA

[3]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

[4]Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, UK

[5]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

[6]Department of Oncology, School of Medicine, Johns Hopkins University, USA

*These authors contributed equally to the manuscript.

**These authors contributed equally to the manuscript.

Running title: Breast cancer risk prediction models: validation and future risk stratification

Keywords: breast cancer, risk prediction, risk stratification, validity, prevention, screening

Corresponding author:

Prof. Nilanjan Chatterjee

Department of Biostatistics

Bloomberg School of Public Health

Johns Hopkins University, USA

(410) 955-3067

nchatte2@jhu.edu

Abstract

Background: External validation of risk models is critical for risk stratified breast cancer prevention. We used the Individualized Coherent Absolute Risk Estimation (iCARE) as a flexible tool for risk model development, comparative model validation, and to make projections for population risk stratification.

Methods: Performance of two recently developed models, iCARE-BPC3 and iCARE-Lit, were compared with two established models (BCRAT, IBIS) based on classical risk factors in a UK-based cohort of 64,874 White non-Hispanic women (863 cases) aged 35-74 years. Risk projections in a target population of US White non-Hispanic women aged 50-70 years assessed potential improvements in risk stratification by adding mammographic breast density (MD) and polygenic risk score (PRS).

Results: The best calibrated models were iCARE-Lit (expected to observed number of cases (E/O)=0.98 (95% confidence interval [CI]=0.87 to 1.11)) for women younger than 50 years; and iCARE-BPC3 (E/O=1.00 (0.93 to 1.09)) for women 50 years or older. Risk projections using iCARE-BPC3 indicated classical risk factors can identify ~500,000 women at moderate to high risk (>3% five-year risk) in the target population. Addition of MD and a 313-variant PRS is expected to increase this to ~3.5 million, and among them, ~153,000 invasive breast cancer cases are expected within five years.

Conclusions: iCARE models based on classical risk factors perform similarly or better than BCRAT or IBIS in White non-Hispanic women. Addition of MD and PRS can lead to substantial improvements in risk stratification. However, these integrated models require independent prospective validation before broad clinical applications.

Breast cancer risk prediction models are used in clinical and research settings to identify women at elevated risk of disease who could benefit from preventive therapies, enhanced screening, or be eligible to participate in prevention trials. Continuing updates of risk models incorporating additional risk factors will potentially improve our ability to identify such women (1).

Independent prospective validation of models is critical to determine their accuracy of prediction, robustness and potential for clinical application. BCRAT and IBIS are established models that originally included hormonal and environmental risk factors and are currently used for clinical and research applications (2). BCRAT has been extensively validated, generally showing good calibration but low risk discrimination (2-4). IBIS performed better than BCRAT in average- to high-risk populations (5,6). Although addition of mammographic breast density (MD) (7-11) or polygenetic risk scores (PRS) (12-20) can lead to improved risk stratification, prospective evaluation of the accuracy of absolute risk predictions from models incorporating PRS is currently lacking.

Risk prediction models should be dynamic and flexible in their ability to incorporate additional risk factors and context-specific incidence rates. However, developing and validating a comprehensive model is challenging due to all relevant risk factors not being typically measured in a single study, and require novel methods for data integration from multiple epidemiologic studies (21-23). Our recently developed Individualized Coherent Absolute Risk Estimation (iCARE) software, implements a flexible approach to build absolute risk models for a population combining information on relative risk estimates, age-specific incidence/mortality rates and risk factor distributions from multiple data sources (24-26). It includes advanced features to account for missing risk factors using internal imputation and a validation component to facilitate comparative model validation across multiple cohorts using uniform methodology (26).

We previously used iCARE to develop a breast cancer risk model using relative risks from a multivariate regression based on eight prospective cohorts of women aged 50 years or older (iCARE-BPC3) (24). Here, we develop an updated version of the synthetic model, described in Garcia-Closas et al. (1), using relative risks from published literature (iCARE-Lit). A literature-based model, while

requiring more assumptions, can include comprehensive sets of risk factors that may not all be measured in one study.

The current study aims to compare the performances of the iCARE models, BCRAT, and IBIS based on classical risk factors (i.e., questionnaire-based risk factors like menstrual, reproductive, hormonal, and lifestyle risk factors) in the UK-based Generations Study (GS). Additionally, the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, used to develop iCARE-BPC3, provided further evaluation of the other models. Risk projections in a target population were estimated based on classical risk factors, and after addition of MD (27) and PRS.

Materials and Methods

Study populations

Primary analyses were performed in a population of 113,211 women aged 16-102 years at enrollment (2003-2012) from the UK-based GS. Further validation of the iCARE-Lit model was performed in 78,214 women aged 50-75 years at enrollment (1993-2001) from the US-based PLCO. Exclusion criteria included history of breast cancer, non-white or unknown ethnicity, no genetic consent or DNA source, entry age below 35 or above 75 years, presence of first- or second-degree relative in study (GS only) and subjects with unconfirmable report of breast cancer (PLCO only). The final analytic samples from the GS and PLCO were 64,874 (863 cases within five years) and 48,279 (1,008 cases within five years), respectively (**Supplementary Figure 1**). As PLCO was used for the development of iCARE-BPC3 (24), it was only used for validating other models. **Supplementary Table 1** shows the risk factor distributions in both cohorts.

Breast Cancer Risk Model Validation and Risk Projection

The **Supplementary Tables 2-4** provide detailed descriptions of the iCARE-based models, BCRAT and IBIS**.** All models incorporate information on marginal disease incidence rates

(**Supplementary Figures 2A, 2B**) and account for competing mortality using mortality rates, both available from population-based registries (26,28,29). The incidence rates were used to calibrate the average predicted risk to the national breast cancer risk (28,29). iCARE implements this step using an additional individual-level reference dataset of risk factors representing the underlying population.

For evaluating calibration, we categorized individuals based on deciles of both the five-year absolute risks that incorporates the variation of age and the relative risk score (i.e., sum of log relative risks multiplied by risk factors) that does not include age. The predicted and observed risks across risk categories were compared using expected-to-observed (E/O) ratio, calibration slope and intercept. Model discrimination was assessed using area under the curve (AUC) statistics based on both five-year absolute risk and the relative risk score (**Supplementary Methods**).

Risk projections of invasive breast cancer were estimated among US White non-Hispanic women aged 50-70 years using the best calibrated model based on classical risk factors in that group. We also evaluated the net benefit (30-32) of this model for high-risk decisions in that population (**Supplementary Methods**). We explored potential improvements in risk stratification and net benefit with addition of PRS and MD. Apart from the 313-SNP PRS (**Supplementary Table 5**) (33-35), we considered an "improved" PRS incorporating the fraction of additional heritability attributable to common variants, and a "best" PRS incorporating all of the common variant heritability (33,35-37). Theoretical AUC was computed using a normal approximation of the relative risk scores for different combinations: classical risk factors only, PRS only, MD only, and a combined model with all risk factors (25,37,38). Moreover, we considered two high-risk thresholds: 3% corresponding to US Preventive Services Task Force (USPSTF) recommendation for risk-lowering drugs and 6% used by the WISDOM trial as a cutoff for very high risk (39,40); and two low risk thresholds: 0.6% and 1.3%, which are average five-year risks of US women aged 40 and 50 years, respectively. We estimated numbers of women and future cases identified at the extremes of the risk distribution based on the above thresholds (**Supplementary Methods**) (24,26).

Results

Breast Cancer Risk Model Validation

Among women younger than 50 years, all models showed good calibration of relative risk (**Figure 1, Table 1, Supplementary Figure 3A**). Absolute risk was best calibrated for iCARE-Lit (E/O=0.98, 95%CI=0.87 to 1.11) with AUC=65.4 (95%CI=62.1 to 68.7) (**Table 1**). BCRAT tended to underestimate (E/O=0.85, 95%CI=0.75 to 0.95) and IBIS to overestimate (E/O=1.14, 95%CI=1.01 to 1.29) absolute risk.

Among women aged 50 years or older, iCARE-BPC3 showed good calibration of absolute and relative risk with E/O=1.00 (95%CI=0.93 to 1.09) and AUC=60.2 (95%CI=58.0 to 62.4). iCARE-Lit showed good calibration of relative risk but overestimation of absolute risk (E/O=1.13, 95%CI=1.04 to 1.22) (**Table 1, Figure 2, Supplementary Figure 4A**). BCRAT and IBIS showed miscalibration of both absolute and relative risk. BCRAT tended to show underestimation in low-risk deciles and overestimation in the high-risk decile. IBIS (E/O=1.13, 95%CI=1.05 to 1.23) showed similar extent of overall miscalibration as iCARE-Lit, and greater miscalibration in the high-risk deciles (**Table 1, Figure 2, Supplementary Figure 4A**).

In PLCO, iCARE-Lit produced similar overestimation of five-year absolute risk as in the GS for women aged 50 years or older. Both BCRAT and IBIS underestimated absolute risk (**Supplementary Table 6, Supplementary Figure 5A**). In both cohorts, discriminatory accuracy was lower when AUC was defined using the relative risk score, as opposed to absolute risk. (**Supplementary Figures 3B, 4B, 5B**).

Breast Cancer Risk Projections

Figure 3 shows five-year absolute risk projections in a target population of White non-Hispanic US women aged 50-70 years (~30 million according to 2016 US Census). MD and 313-SNP

PRS alone had higher AUCs compared to classical risk factors (based on iCARE-BPC3). An integrated model with classical risk factors, MD and PRS had the highest AUC of 68.3.

The classical risk factors could identify approximately 4.1 million women, representing 13.8% of the target population, at low risk (<1.13%, corresponding to the average five-year risk for 50-year-old US women) of invasive breast cancer and 40,516 (8.2% of all cases) are expected to develop the disease within five years (**Figure 4B, Supplementary Table 7**). Integrating classical risk factors with MD and 313-SNP PRS is expected to increase the number of women to 12 million, and around 89,000 (17.7% of all cases) would be expected to develop the disease within five years. In the moderate- to high-risk group (>3% five-year risk threshold based on USPSTF recommendation for risk-reducing therapies (40)), approximately 500,000 women, representing 1.7% of this population, could be identified based on classical risk factors, including approximately nearly 17,000 (3.4% of all cases) expected to develop the disease within five years (**Figure 4C, Supplementary Table 7**). Integrating with MD and 313-SNP PRS increases the number of women identified to 3.5 million and among them, approximately 153,000 (~30% of all cases) would be expected to develop disease within five years.

We projected that doubling the size of current breast cancer GWAS (to around 300,000 cases and 300,000 controls) would yield additional discoveries and an "improved" PRS with AUC=69.1 (**Supplementary Table 7**). An integrated model with improved PRS could identify approximately 14 million women at low risk, and approximately 92,000 (~18% of all cases) would be expected to develop invasive breast cancer within five years (**Figure 4B, Supplementary Table 7**). In the moderate- to high-risk group, we could identify close to 4.2 million women, with approximately 207,000 (~40% of all cases) expected to develop the disease within five years. This is close to the risk stratification attained by the "best" theoretical PRS explaining 100% of the variability of polygenic risk from common variants (**Figure 4C, Supplementary Table 7**). The relative increase in the numbers of women and cases identified due to incorporation of the improved PRS also increases as the risk

threshold becomes more extreme at either end of the distribution **(Figures 4A-4D, Supplementary Table 7)**.

Evaluation of Net Benefit

We evaluated the theoretical net benefit for high-risk decisions based on the iCARE-BPC3 model and its extensions after addition of PRS and MD in the US population of White non-Hispanic women aged 50-70 years (**Figure 5**). At a 3% five-year risk threshold used for recommendation for risk-lowering medications (40) there is virtually no net benefit for a model with classical risk factors alone, while the integrated models with the addition of PRS and MD show some net benefit for these women. However, none of the models show net benefit for risk-reducing interventions on women at the highest risk threshold (i.e., above 6%).

Discussion

In this comparative analysis using data from a large population-based cohort, we showed that iCARE-based absolute risk models for invasive breast cancer with classical risk factors are similarly or better calibrated than previous models evaluated here, and that the addition of MD and PRS to classical risk factors can substantially improve risk stratification in the population.

Among women younger than 50 years, we found no substantial evidence of miscalibration of the relative risk for any of the models evaluated; however, we found some evidence for miscalibration of the five-year absolute risk for all models except iCARE-Lit. This illustrates the challenges of validating models for absolute risk since in addition to relative risk information, it requires information on population-based incidence rates and distribution of risk factors, ideally from the same time period as the validation study.

Among women aged 50 years or older, we found no evidence for miscalibration of iCARE-BPC3 in terms of relative or absolute risk, while the other models overestimated absolute risk for women in the highest risk category. While relative risk is reasonably well calibrated for both iCARE-

based models, the better absolute risk calibration for iCARE-BPC3 compared to iCARE-Lit could be due to differences in specification of risk factors (i.e., finer versus coarser categories of continuous risk factors).

The IBIS and iCARE models include more information on classical risk factors than BCRAT (28,29), which has lower discriminatory accuracy. IBIS incorporates detailed family history information, considered important to identify high-risk women with extensive family history. However, comparisons of risk stratification across models in our study were limited by the miscalibration of BCRAT, IBIS and iCARE-Lit for women aged 50 years or older, particularly in high-risk deciles.

BCRAT has been extensively evaluated and is currently recommended for predicting breast cancer risk for US women undergoing mammographic screening (41,42). While some studies found no evidence for miscalibration of BCRAT (43-46), others reported underestimation (47-49) or overestimation (46,50-52) of risk. Some studies reported improved calibration when using incidence and mortality rates from the same country and time period (48,49,52). Sensitivity analyses using rates closer to our validation population also indicated slight improvement in calibration (data not shown). Miscalibration may be due to model misspecification or differences in the risk factor distribution between the cohort and underlying population.

Two validation studies in high-risk populations in the US and UK found no evidence for miscalibration of IBIS (5,53,54). IBIS 10-year risk predictions have been found to be well calibrated in family-based studies including average-to-high risk women, in contrast to underestimation of risk by BCRAT in these populations (5,6). Both IBIS and BCRAT showed good absolute risk calibration in an Australian population of average-risk women (16). A recent prospective evaluation in a US-based integrated healthcare system showed good calibration of IBIS 10-year absolute risks overall, but ~20% overestimation in the highest risk decile (9). The current literature and our findings highlight the importance of external validation of risk models using multiple prospective studies to evaluate robustness of model performance across populations. Of note, validation studies often lack

adequate size and rigorous methodology, making comparisons across studies difficult (55). There is a need for further robust validation in large studies to identify models adequate for clinical decision making (56).

Our study has several strengths. GS is a recent population-based cohort including women with a wide age range. Moreover, the validation results of the iCARE-Lit model in PLCO further supported our overall conclusions. We evaluated model calibration overall and stratified by levels of risk. The latter is important as accurate classification of subjects at the extremes of risk is most relevant for risk-based screening. Second, we assessed model calibration by deciles of expected absolute risk and by the relative risk score. The former is commonly used in validation studies (5,43-45,47,52,53) as absolute risk is the relevant measure for clinical or public health applications. However, strong dependence of absolute risk on age makes the differences in model performance due to other risk factors less evident than comparisons using the relative risk score, which does not include age. We evaluated model calibration and discrimination with and without accounting for age. Most (5,44,47-50,53,54) but not all (43,45,52) previous validation studies of BCRAT and IBIS assessed model discrimination accounting for age. Such model discrimination statistics (e.g., AUC) evaluated in a validation cohort may differ from those in the target population owing to differences in risk factor distributions. Additionally, our results showed that small changes in overall measures (e.g., AUC) derived from additional risk factors can result in substantial changes in the number of women at the extremes of risk distribution.

Limitations of our analyses include that not all risk factors were available in the validation cohorts. Additional validation of iCARE models in cohorts within representative healthcare systems is desirable for more robust model evaluation. Moreover, we only evaluated short-term risk predictions, assuming risk factors remained constant over the prediction period. Validation of long-term risk will require further follow-up or additional studies, preferably accounting for time-varying risk factors and time-dependent associations. Our current model development and validation efforts focus on predicting risk of breast cancer for White non-Hispanic women. Our ongoing work will

extend the models to non-White populations and include extensive classical risk factor information and an improved PRS (57,58).

For a given high-risk clinical decision, a well-calibrated model providing wider risk stratification is likely to have greater clinical utility. We have shown this using theoretical net benefit analyses in a target population; however, further assessment of the clinical utility of models will require identifying risk thresholds explicitly informed by benefits and costs of a specific intervention (20,59,60). Risk projections based on the integrated model assumed that classical risk factors, MD, and PRS act multiplicatively on disease risk. We accounted for known dependencies between classical risk factors, and previous studies support multiplicative effects of classical risk factors or MD with PRS on disease risk (24,61). Risk projections based on models with MD accounted for its dependence on age and BMI, but not on the other risk factors in the model that have weaker associations with MD (62). Thus, these should only be considered as projections, and such integrated models require independent prospective validation prior to consideration for clinical use. We derived PRS based on SNP odds ratios from genetic discovery studies (33,34). This may result in overestimation of risk stratification; however, based on previous assessment (63) this bias is likely small.

Updates to BCRAT and IBIS have added MD (7-11) and PRS (12-18) to the original models. However, only the addition of MD to IBIS has been prospectively validated and it showed some overestimation in the high-risk categories (9). Addition of an 18-SNP PRS to IBIS was shown to increase risk discrimination in a UK-based screening cohort, though accuracy of prediction was not prospectively evaluated (18). The Breast Cancer Surveillance Consortium risk model, which includes BI-RADS MD and a 76-SNP PRS, has also not been validated prospectively (64).

Further improvements in risk stratification could be achieved by incorporating additional risk factors and heterogeneity in risk factor associations by breast cancer subtypes. Ultimately, it is desirable to develop a comprehensive model, robustly validated in multiple populations with disparate risk factor information, applicable in populations with a wide range of underlying risk and

provides reliable risk estimates based on subsets of risk factors depending on the clinical application (e.g., risk assessment before or after mammography). The iCARE methodology facilitates this by providing a flexible risk modeling and validation tool with capabilities of handling missing risk factors.

In conclusion, we have demonstrated that iCARE models based on classical risk factors perform similarly and, in some cases, better than established models for five-year risk predictions of invasive breast cancer. Based on our projections, substantial improvements in risk stratification can be achieved with addition of MD and PRS to classical risk factors. Such integrated risk models require prospective empirical validation before broad clinical or research applications.

References

1. Garcia-Closas M, Gunsoy NB,Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. Journal of the National Cancer Institute. 2014;106(11).
2. Cintolo-Gonzalez JA, Braun D, Blackford AL, et al. Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. Breast cancer research and treatment. 2017;164(2):263-284.
3. Meads C, Ahmed I,Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast cancer research and treatment. 2012;132(2):365-377.
4. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V,Thakkinstian A. Risk prediction models of breast cancer: a systematic review of model performances. Breast cancer research and treatment. 2012;133(1):1-10.
5. Quante AS, Whittemore AS, Shriver T, Strauch K,Terry MB. Breast cancer risk assessment across the risk continuum: genetic and nongenetic risk factors contributing to differential model performance. Breast cancer research : BCR. 2012;14(6):R144.
6. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. The Lancet Oncology. 2019;20(4):504-517.
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 research and treatment. 2005;94(2):115-122.
8. Chen J, Pee D, Ayyagari R, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. Journal of the National Cancer Institute. 2006;98(17):1215-1226.
9. 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:e180174.

10. Brentnall AR, Harkness EF, Astley SM, 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 research : BCR. 2015;17(1):147.
11. Warwick J, Birke H, Stone 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 research : BCR. 2014;16(5):451.
12. Allman R, Dite GS, Hopper JL, et al. SNPs and breast cancer risk prediction for African American and Hispanic women. Breast cancer research and treatment. 2015;154(3):583-589.
13. Cuzick J, Brentnall AR, Segal C, et al. Impact of a Panel of 88 Single Nucleotide Polymorphisms on the Risk of Breast Cancer in High-Risk Women: Results From Two Randomized Tamoxifen Prevention Trials. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2017;35(7):743-750.
14. Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M,Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. Journal of the National Cancer Institute. 2010;102(21):1618-1627.
15. Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. Journal of the National Cancer Institute. 2009;101(13):959-963.
16. Dite GS, MacInnis RJ, Bickerstaffe A, et al. Breast Cancer Risk Prediction Using Clinical Models and 77 Independent Risk-Associated SNPs for Women Aged Under 50 Years: Australian Breast Cancer Family Registry. Cancer Epidemiol Biomarkers Prev. 2016;25(2):359-365.
17. Evans DG, Brentnall A, Byers H, et al. The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case-control study. J Med Genet. 2017;54(2):111-113.
18. van Veen EM, Brentnall AR, Byers H, et al. Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. JAMA Oncol. 2018;4(4):476-482.
19. Torkamani A, Wineinger NE,Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018.
20. Pashayan N, Morris S, Gilbert FJ,Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. JAMA Oncol. 2018.
21. Chatterjee N, Chen YH, Maas P,Carroll RJ. Constrained Maximum Likelihood Estimation for Model Calibration Using Summary-level Information from External Big Data Sources. J Am Stat Assoc. 2016;111(513):107-117.
22. Grill S, Ankerst DP, Gail MH, Chatterjee N,Pfeiffer RM. Comparison of approaches for incorporating new information into existing risk prediction models. Stat Med. 2017;36(7):1134-1156.
23. Tang R, Kundu P,Chatterjee N. Generalized Meta-Analysis for Multivariate Regression Models Across Studies with Disparate Covariate Information. arXiv:170803818 [statME]. 2017.
24. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol. 2016;2(10):1295-1302.
25. Chatterjee N, Shi J,Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet. 2016;17(7):392-406.
26. Pal Choudhury P, Maas P, Wilcox A, et al. iCARE: An R Package to Build, Validate and Apply Absolute Risk Models. bioRxiv. 2018. https://doi.org/10.1101/079954.
27. 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):1159-1169.

28. Gail MH, Brinton LA, Byar DP, 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):1879-1886.
29. Tyrer J, Duffy SW,Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004;23(7):1111-1130.
30. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-138.
31. Rousson V,Zumbrunn T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. BMC Med Inform Decis Mak. 2011;11:45.
32. Zhang Z, Rousson V, Lee WC, et al. Decision curve analysis: a technical note. Ann Transl Med. 2018;6(15):308.
33. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92-94.
34. Milne RL, Kuchenbaecker KB, Michailidou K, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet. 2017;49(12):1767-1778.
35. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet. 2018.
36. Zhang Y, Qi G, Park JH,Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat Genet. 2018;50(9):1318-1326.
37. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ,Park JH. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet. 2013;45(4):400-405, 405e401-403.
38. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF,Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet. 2002;31(1):33-36.
39. Shieh Y, Eklund M, Madlensky L, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. Journal of the National Cancer Institute. 2017;109(5).
40. Final Recommendation Statement: Breast Cancer: Medications for Risk Reduction. U.S. Preventive Services Task Force. https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/breast-cancer-medications-for-risk-reduction. Published 2016. Accessed December 13, 2018.
41. Practice Bulletin Number 179: Breast Cancer Risk Assessment and Screening in Average-Risk Women. Obstet Gynecol. 2017;130(1):e1-e16.
42. NCCN Clinical Practice Guidelines in Oncology: Breast Cancer Screening and Diagnosis, Version 1.2018. In: National Comprehensive Cancer Network (www.nccn.org); 2018.
43. Decarli A, Calza S, Masala G, Specchia C, Palli D,Gail MH. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence-European Prospective Investigation Into Cancer and Nutrition cohort. Journal of the National Cancer Institute. 2006;98(23):1686-1693.
44. Arrospide A, Forne C, Rue M, Tora N, Mar J,Bare M. An assessment of existing models for individualized breast cancer risk estimation in a screening program in Spain. BMC Cancer. 2013;13:587.
45. 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. Annals of internal medicine. 2008;148(5):337-347.
46. Nickson C, Procopio P, Velentzis LS, et al. Prospective validation of the NCI Breast Cancer Risk Assessment Tool (Gail Model) on 40,000 Australian women. Breast cancer research : BCR. 2018;20(1):155.

47. Chlebowski RT, Anderson GL, Lane DS, et al. Predicting risk of breast cancer in postmenopausal women by hormone receptor status. Journal of the National Cancer Institute. 2007;99(22):1695-1705.
48. Schonfeld SJ, Pee D, Greenlee RT, et al. Effect of changing breast cancer incidence rates on the calibration of the Gail model. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(14):2411-2417.
49. Rockhill B, Spiegelman D, Byrne C, Hunter DJ,Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. Journal of the National Cancer Institute. 2001;93(5):358-366.
50. Banegas MP, Gail MH, LaCroix A, et al. Evaluating breast cancer risk projections for Hispanic women. Breast cancer research and treatment. 2012;132(1):347-353.
51. Spiegelman D, Colditz GA, Hunter D,Hertzmark E. Validation of the Gail et al. model for predicting individual breast cancer risk. Journal of the National Cancer Institute. 1994;86(8):600-607.
52. Pastor-Barriuso R, Ascunce N, Ederra M, et al. Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. Breast cancer research and treatment. 2013;138(1):249-259.
53. Powell M, Jamshidian F, Cheyne K, Nititham J, Prebil LA,Ereman R. Assessing breast cancer risk models in Marin County, a population with high rates of delayed childbirth. Clin Breast Cancer. 2014;14(3):212-220.e211.
54. Amir E, Evans DG, Shenton A, et al. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. J Med Genet. 2003;40(11):807-814.
55. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC medical research methodology. 2014;14:40.
56. Garcia-Closas M,Chatterjee N. Assessment of breast cancer risk: which tools to use? The Lancet Oncology. 2019;20(4):463-464.
57. Confluence Project. NCI Division of Cancer Epidemiology and Genetics. https://dceg.cancer.gov/research/cancer-types/breast-cancer/confluence-project. Accessed December 21, 2018.
58. NCI Cohort Consortium. National Cancer Institute. https://epi.grants.cancer.gov/Consortia/cohort.html. Accessed January 27, 2019.
59. Mandelblatt JS, Stout NK, Schechter CB, et al. Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies. Annals of internal medicine. 2016;164(4):215-225.
60. Trentham-Dietz A, Kerlikowske K, Stout NK, et al. Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes. Annals of internal medicine. 2016;165(10):700-712.
61. Rudolph A, Song M, Brook MN, et al. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. International journal of epidemiology. 2018.
62. Alexeeff SE, Odo NU, McBride R, et al. Reproductive Factors and Mammographic Density: Associations Among 24,840 Women and Comparison of Studies Using Digitized Film-Screen Mammography and Full-Field Digital Mammography. American journal of epidemiology. 2019.
63. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. Journal of the National Cancer Institute. 2015;107(5).
64. Vachon CM, Pankratz VS, Scott CG, et al. The contributions of breast density and common genetic variation to breast cancer risk. Journal of the National Cancer Institute. 2015;107(5).

Funding

This work was supported by the Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1602-34530), Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics (Z01CP010119), and the European Union's Horizon 2020 research and innovation programme under grant agreements No 633784 (B-CAST).

The Generations Study would like to thank Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study, and the Study participants, Study staff, and the doctors, nurses and other health care staff and data providers who have contributed to the Study. The ICR acknowledge NHS funding to the NIHR Biomedical Research Centre.

Notes

The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.
The authors have no conflicts of interest related to this work.

We would like to thank Adam R. Brentnall and Jack Cuzick for providing the IBIS Risk Evaluator software and for their comments to a draft of the manuscript. We would like to thank Celine Vachon for providing unpublished information on the association of mammographic breast density with invasive breast cancer and population distribution of mammographic density. This information was used in the risk projection calculations.

Tables

Table 1. Ratios of expected to observed five-year absolute risk for the breast cancer risk prediction models validated using the GS*

Age group	Model	AUC (95% CI)	Overall			Top risk decile		
			O% (95% CI)	E%	E/O ratio (95% CI)	O% (95% CI)	E%	E/O ratio (95% CI)
<50 years of age (265 cases, 27,967 non-cases)	iCARE-Lit	65.4 (62.1 to 68.7)	0.94 (0.83 to 1.05)	0.92	0.98 (0.87 to 1.11)	2.22 (1.68 to 2.77)	2.51	1.13 (0.89 to 1.44)
	BCRAT	64.0 (60.6 to 67.4)		0.79	0.85 (0.75 to 0.95)	2.24 (1.70 to 2.79)	1.73	0.77 (0.60 to 0.98)
	IBIS	64.6 (61.3 to 67.9)		1.07	1.14 (1.01 to 1.29)	2.19 (1.65 to 2.73)	2.58	1.18 (0.92 to 1.51)
≥50 years of age (598 cases, 36,044 non-cases)	iCARE-Lit	62.2 (60.0 to 64.5)	1.63 (1.50 to 1.76)	1.84	1.13 (1.04 to 1.22)	3.20 (2.62 to 3.77)	3.91	1.22 (1.02 to 1.46)
	iCARE-BPC3	60.2 (58.0 to 62.4)		1.64	1.00 (0.93 to 1.09)	2.63 (2.12 to 3.15)	2.85	1.08 (0.89 to 1.32)
	BCRAT	58.2 (55.8 to 60.5)		1.56	0.95 (0.88 to 1.03)	2.73 (2.20 to 3.26)	3.27	1.20 (0.99 to 1.46)
	IBIS	61.4 (59.2 to 63.6)		1.85	1.13 (1.05 to 1.23)	3.07 (2.51 to 3.63)	3.98	1.30 (1.08 to 1.56)

* AUC = area under the curve; GS = Generations Study; CI = confidence interval; E = expected absolute risk; O = observed absolute risk. The AUCs reported in Table 1 are defined based on absolute risk and incorporate the variation due to age. The AUCs (95%CI) based on the relative risk score, which do not incorporate variation of age, are as follows: (i) for women younger than 50 years, iCARE-Lit: 58.8 (55.3 to 62.3), BCRAT: 54.6 (50.9 to 58.4), IBIS: 57.0 (53.4 to 60.6); (ii) for women 50 years or older, iCARE-Lit: 60.3 (58.0 to 62.6), iCARE-BPC3: 57.7 (55.4 to 60.0), BCRAT: 52.2 (49.6 to 54.7), IBIS: 60.2 (57.9 to 62.5).

Figure titles and legends

Figure 1. Absolute risk calibration of breast cancer risk prediction models in the GS cohort among women less than 50 years of age. The risk categories are based on absolute risk. The backgrounds of the plots are shaded to indicate the absolute risk threshold categories: ≤0.6% in pink, >0.6% to ≤1.13% in green, >1.13% to 3% in blue, and >3% in orange. The 0.6% and 1.13% thresholds correspond to the average five-year risk for US women aged 40 years and 50 years, respectively. The 3% threshold is used by the United States Preventive Services Task Force for recommending risk-lowering drugs and 6% is used by the WISDOM trial as a threshold for very high risk. E/O = expected-to-observed, GS = Generations Study.

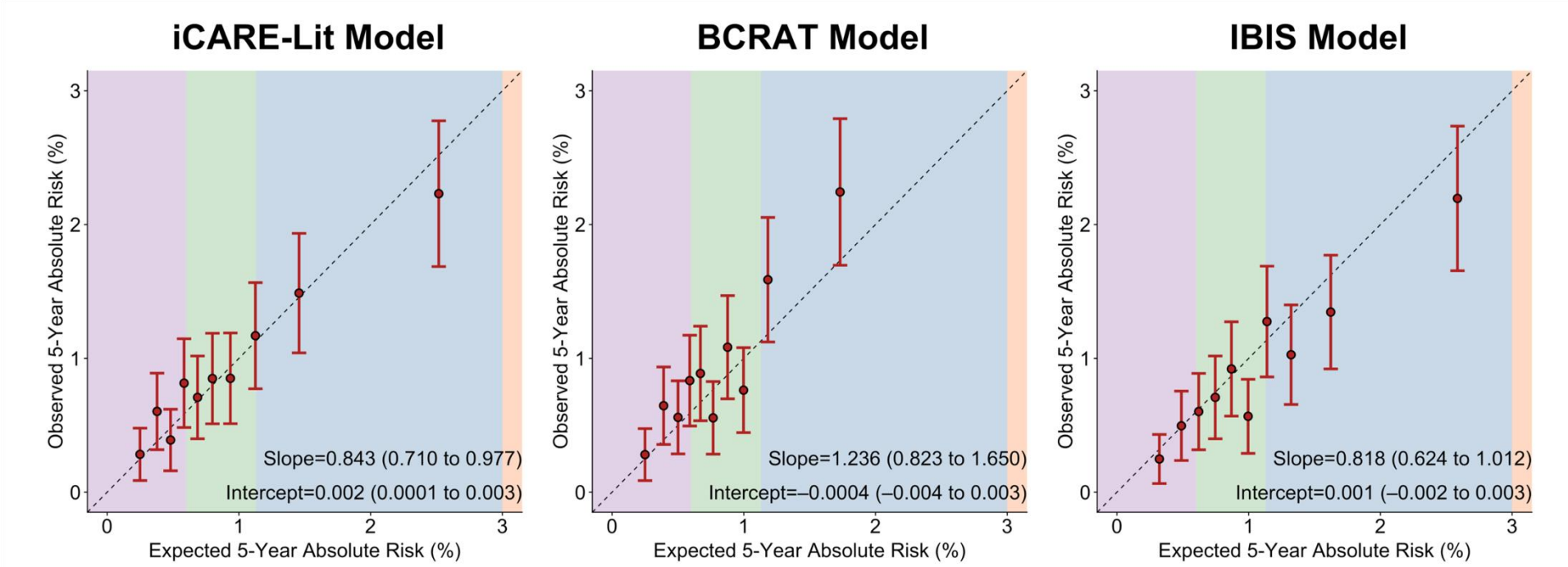
Figure 2. Absolute risk calibration of breast cancer risk prediction models in the GS cohort among women 50 years of age or greater. The risk categories are based on absolute risk. The backgrounds of the plots are shaded to indicate the absolute risk threshold categories: ≤0.6% in pink, >0.6% to ≤1.13% in green, >1.13% to 3% in blue, and >3% in orange. The 0.6% and 1.13% thresholds correspond to the average five-year risk for US women aged 40 years and 50 years, respectively. The 3% threshold is used by the United States Preventive Services Task Force for recommending risk-lowering drugs and 6% is used by the WISDOM trial as a threshold for very high risk. E/O = expected-to-observed, GS = Generations Study.

Figure 3. Five-year absolute risk projection for the general US population of White non-Hispanic women, ages 50-70 years. The classical risk factors correspond to the iCARE-BPC3 model. Classical risk factors include age at menarche, age at menopause, parity, age at first birth, height, alcohol intake, breast cancer family history, smoking status, BMI, current HRT use, and ever HRT type. The projected AUCs reported are based on the relative risk score in that population and do not incorporate variation due to age. AUC = area under the curve, MD = mammographic breast density, PRS = polygenic risk score, SNP = single nucleotide polymorphism.

Figure 4. White non-Hispanic women aged 50-70 years in the US population expected to be identified at elevated risk of breast cancer according to different risk thresholds, and the incident cases of invasive breast cancer who are expected to occur in these groups within a five-year interval. The expected number of subjects is calculated using mid-2016 population estimates (n=30,030,821) from the US Census Bureau and the number of cases is calculated using the average predicted five-year risk and the 2015 invasive breast cancer incidence rates from SEER. The 0.6% and 1.13% thresholds correspond to the average five-year risk for US women aged 40 years and 50 years, respectively. The 3% threshold is used by the United States Preventive Services Task Force for recommending risk-lowering drugs and 6% is used by the WISDOM trial as a threshold for very high risk. The projected AUCs reported are based on the relative risk score in that population and do not incorporate variation due to age. AUC = area under the curve, MD = mammographic breast density, PRS = polygenic risk score, SNP = single nucleotide polymorphism.

Figure 5. Projected net benefit of identifying White non-Hispanic women aged 50-70 years with predicted absolute risk above a range of thresholds, with vertical lines representing 3% and 6% thresholds. The 3% threshold is used by the United States Preventive Services Task Force for recommending risk-lowering drugs and 6% is used by the WISDOM trial as a threshold for very high risk. The classical risk factors correspond to the iCARE-BPC3 model. Classical risk factors include age at menarche, age at menopause, parity, age at first birth, height, alcohol intake, breast cancer family history, smoking status, BMI, current HRT use, and ever HRT type. Projections were made under the assumption of perfect calibration and log-normal distribution of risk in the population and using information on the spread (standard deviation) of the risk score from the reference sample and distribution of age in the current US population. MD = mammographic breast density, PRS = polygenic risk score, SNP = single nucleotide polymorphism.

Figure 1

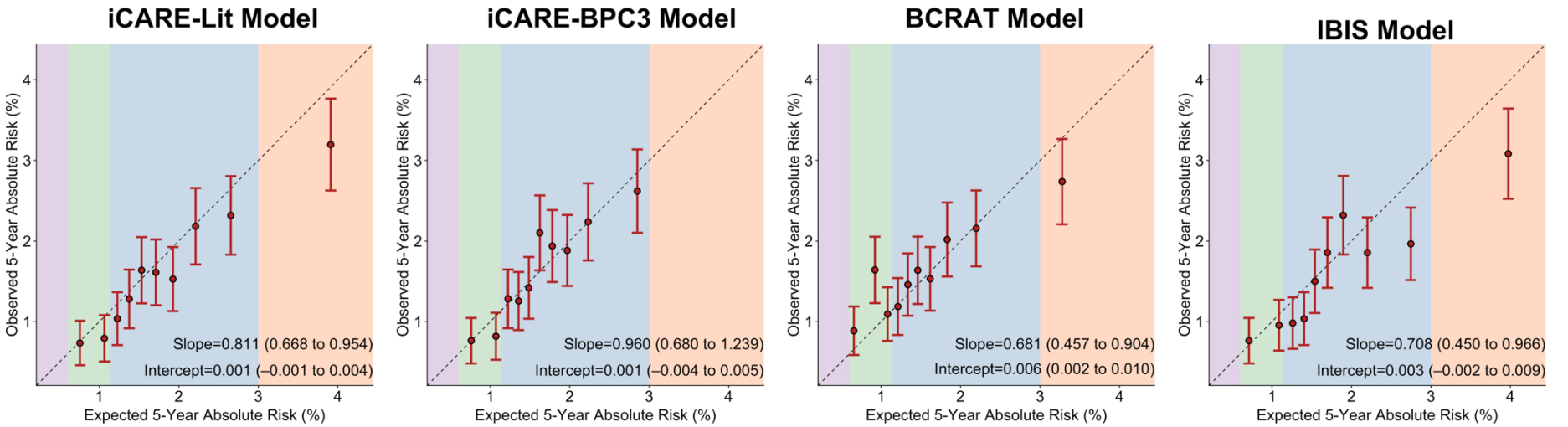


Decile	Total subjects (n)	Cases (n)	E/O ratio (95% CI)
1	2 847	8	0.89 (0.44 to 1.77)
2	2 804	17	0.63 (0.39 to 1.00)
3	2 829	11	1.24 (0.69 to 2.23)
4	2 808	23	0.71 (0.47 to 1.07)
5	2 820	20	0.97 (0.62 to 1.50)
6	2 827	24	0.94 (0.63 to 1.40)
7	2 759	22	1.17 (0.77 to 1.77)
8	2 906	35	0.93 (0.67 to 1.30)
9	2 798	42	0.97 (0.72 to 1.31)
10	2 834	63	1.13 (0.89 to 1.44)
Overall	28 232	265	0.98 (0.87 to 1.11)

Decile	Total subjects (n)	Cases (n)	E/O ratio (95% CI)
1	2 713	7	0.95 (0.46 to 2.00)
2	2 914	18	0.62 (0.39 to 0.99)
3	3 027	18	0.84 (0.53 to 1.33)
4	2 539	19	0.78 (0.50 to 1.23)
5	2 925	28	0.70 (0.48 to 1.01)
6	2 697	15	1.37 (0.83 to 2.27)
7	2 926	31	0.82 (0.58 to 1.17)
8	2 926	22	1.33 (0.88 to 2.01)
9	2 757	44	0.74 (0.55 to 0.99)
10	2 808	63	0.77 (0.60 to 0.98)
Overall	28 232	265	0.85 (0.75 to 0.95)

Decile	Total subjects (n)	Cases (n)	E/O ratio (95% CI)
1	2 820	7	1.29 (0.62 to 2.72)
2	2 844	14	0.99 (0.59 to 1.68)
3	2 800	17	1.02 (0.64 to 1.64)
4	2 833	20	1.06 (0.68 to 1.64)
5	2 835	26	0.95 (0.65 to 1.39)
6	2 737	16	1.70 (1.04 to 2.77)
7	2 887	36	0.91 (0.66 to 1.26)
8	2 794	28	1.32 (0.91 to 1.90)
9	2 848	39	1.18 (0.87 to 1.62)
10	2 834	62	1.18 (0.92 to 1.51)
Overall	28 232	265	1.14 (1.01 to 1.29)

Figure 2



Decile	Total subjects (n)	Cases (n)	E/O ratio (95% CI)
1	3 662	27	1.02 (0.70 to 1.49)
2	3 669	29	1.35 (0.94 to 1.93)
3	3 613	38	1.17 (0.85 to 1.60)
4	3 707	47	1.08 (0.82 to 1.44)
5	3 705	62	0.92 (0.72 to 1.17)
6	3 618	57	1.09 (0.84 to 1.41)
7	3 613	54	1.29 (0.99 to 1.68)
8	3 729	82	1.00 (0.81 to 1.24)
9	3 666	85	1.14 (0.93 to 1.41)
10	3 660	117	1.22 (1.02 to 1.46)
Overall	36 642	598	1.13 (1.04 to 1.22)

Decile	Total subjects (n)	Cases (n)	E/O ratio (95% CI)
1	3 673	28	1.00 (0.69 to 1.45)
2	3 558	29	1.31 (0.91 to 1.89)
3	3 701	47	0.96 (0.72 to 1.28)
4	3 755	47	1.08 (0.82 to 1.44)
5	3 543	49	1.07 (0.81 to 1.42)
6	3 777	80	0.77 (0.62 to 0.95)
7	3 532	70	0.90 (0.71 to 1.13)
8	3 763	70	1.06 (0.84 to 1.33)
9	3 656	81	1.01 (0.81 to 1.25)
10	3 684	97	1.08 (0.89 to 1.32)
Overall	36 642	598	1.00 (0.93 to 1.09)

Decile	Total subjects (n)	Cases (n)	E/O ratio (95% CI)
1	3 724	33	0.74 (0.53 to 1.04)
2	3 422	56	0.56 (0.43 to 0.73)
3	3 982	45	0.95 (0.71 to 1.27)
4	3 616	43	1.02 (0.76 to 1.37)
5	3 463	52	0.89 (0.68 to 1.16)
6	3 857	60	0.94 (0.73 to 1.21)
7	3 325	51	1.05 (0.80 to 1.38)
8	3 936	79	0.91 (0.73 to 1.13)
9	3 651	79	1.01 (0.82 to 1.26)
10	3 666	100	1.20 (0.99 to 1.46)
Overall	36 642	598	0.95 (0.88 to 1.03)

Decile	Total subjects (n)	Cases (n)	E/O ratio (95% CI)
1	3 665	28	0.93 (0.64 to 1.35)
2	3 570	31	1.25 (0.88 to 1.78)
3	3 846	42	1.15 (0.85 to 1.56)
4	3 574	36	1.40 (1.01 to 1.93)
5	3 556	55	0.99 (0.76 to 1.29)
6	3 677	68	0.92 (0.72 to 1.16)
7	3 716	85	0.83 (0.67 to 1.02)
8	3 708	68	1.20 (0.95 to 1.52)
9	3 679	73	1.39 (1.10 to 1.74)
10	3 651	112	1.30 (1.08 to 1.56)
Overall	36 642	598	1.13 (1.05 to 1.23)

Figure 3

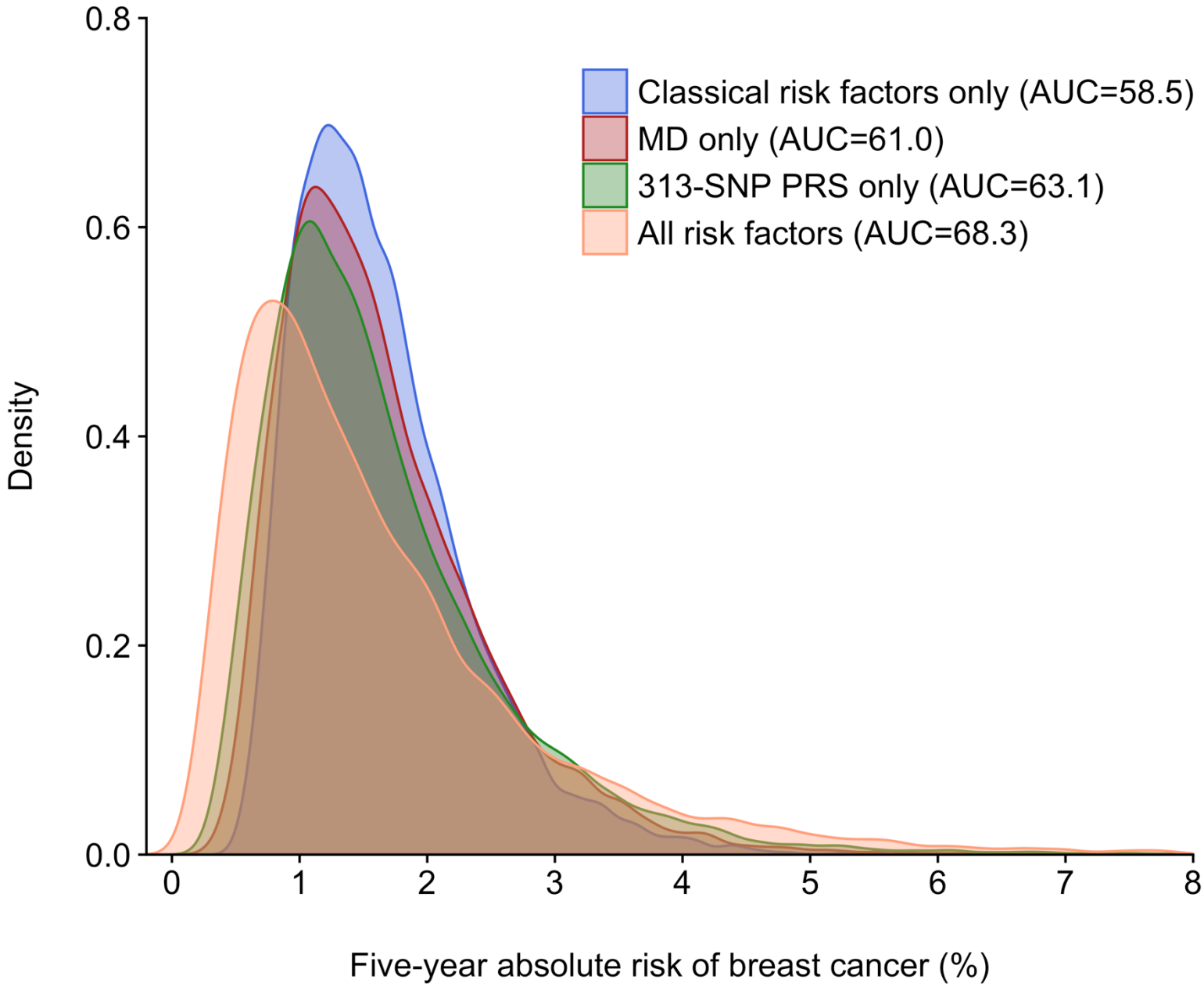
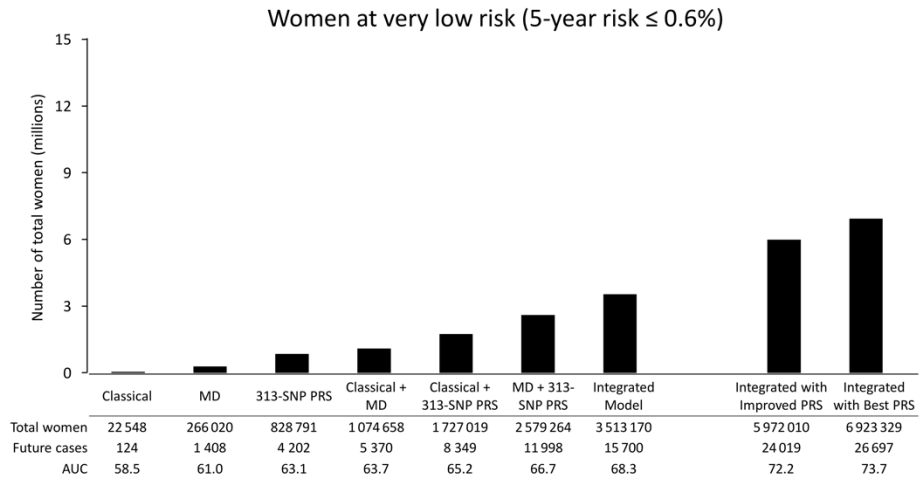
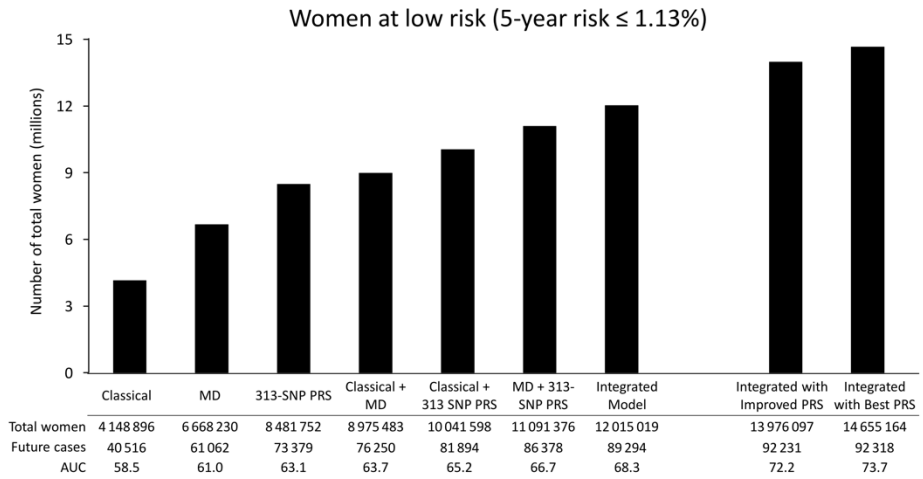

Classical risk factors only (AUC=58.5)
MD only (AUC=61.0)
313-SNP PRS only (AUC=63.1)
All risk factors (AUC=68.3)
Density
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7
8
Five-year absolute risk of breast cancer (%)

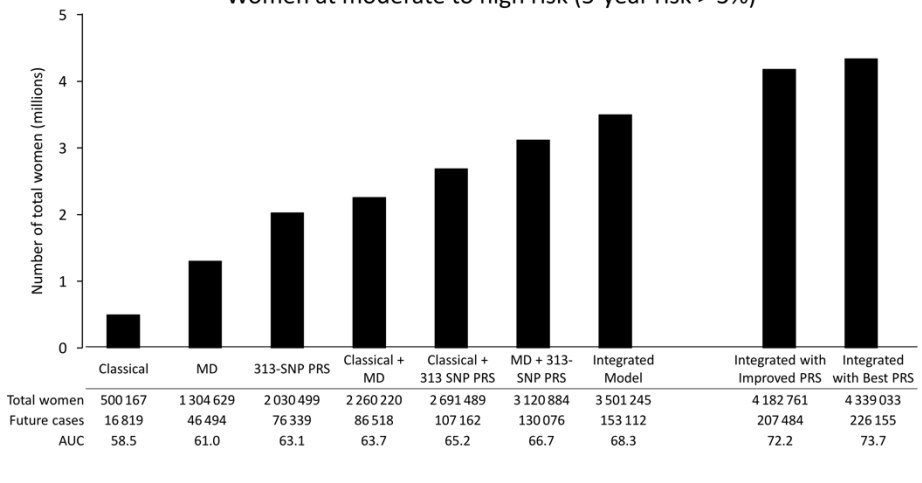
Figure 4



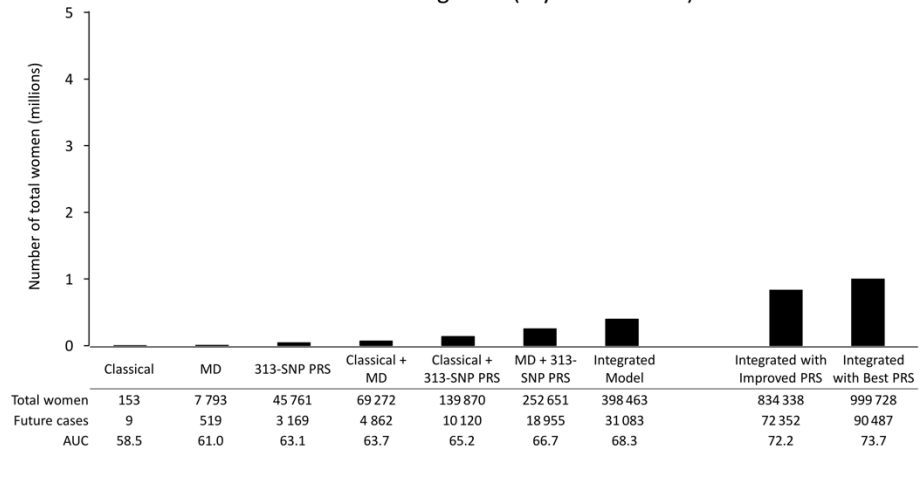
	Classical	MD	313-SNP PRS	Classical + MD	Classical + 313-SNP PRS	MD + 313-SNP PRS	Integrated Model	Integrated with Improved PRS	Integrated with Best PRS
Total women	22 548	266 020	828 791	1 074 658	1 727 019	2 579 264	3 513 170	5 972 010	6 923 329
Future cases	124	1 408	4 202	5 370	8 349	11 998	15 700	24 019	26 697
AUC	58.5	61.0	63.1	63.7	65.2	66.7	68.3	72.2	73.7



	Classical	MD	313-SNP PRS	Classical + MD	Classical + 313 SNP PRS	MD + 313-SNP PRS	Integrated Model	Integrated with Improved PRS	Integrated with Best PRS
Total women	4 148 896	6 668 230	8 481 752	8 975 483	10 041 598	11 091 376	12 015 019	13 976 097	14 655 164
Future cases	40 516	61 062	73 379	76 250	81 894	86 378	89 294	92 231	92 318
AUC	58.5	61.0	63.1	63.7	65.2	66.7	68.3	72.2	73.7

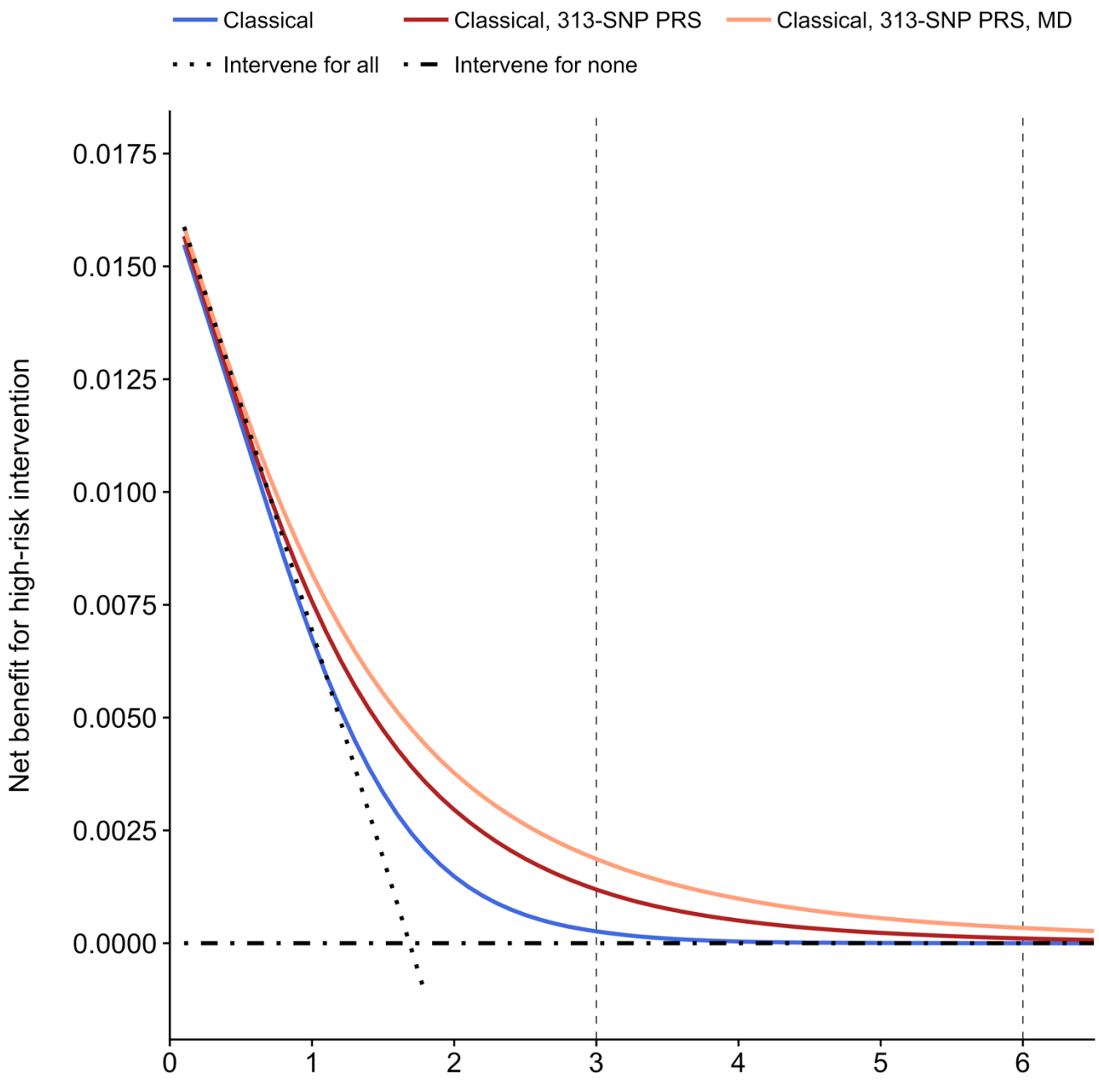


	Classical	MD	313-SNP PRS	Classical + MD	Classical + 313 SNP PRS	MD + 313-SNP PRS	Integrated Model	Integrated with Improved PRS	Integrated with Best PRS
Total women	500 167	1 304 629	2 030 499	2 260 220	2 691 489	3 120 884	3 501 245	4 182 761	4 339 033
Future cases	16 819	46 494	76 339	86 518	107 162	130 076	153 112	207 484	226 155
AUC	58.5	61.0	63.1	63.7	65.2	66.7	68.3	72.2	73.7



	Classical	MD	313-SNP PRS	Classical + MD	Classical + 313-SNP PRS	MD + 313-SNP PRS	Integrated Model	Integrated with Improved PRS	Integrated with Best PRS
Total women	153	7 793	45 761	69 272	139 870	252 651	398 463	834 338	999 728
Future cases	9	519	3 169	4 862	10 120	18 955	31 083	72 352	90 487
AUC	58.5	61.0	63.1	63.7	65.2	66.7	68.3	72.2	73.7

Figure 5


Classical
Classical, 313-SNP PRS
Classical, 313-SNP PRS, MD
Intervene for all
Intervene for none
Net benefit for high-risk intervention
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
0
1
2
3
4
5
6