Genetics in Medicine (2023) 25, 100917

ELSEVIER

Genetics in Medicine
An Official Journal of the ACMG

www.journals.elsevier.com/genetics-in-medicine

ARTICLE

Breast cancer risk stratification using genetic and non-genetic risk assessment tools for 246,142 women in the UK Biobank

Peh Joo Ho[1,2,3], Elaine H. Lim[4], Mikael Hartman[2,3,5], Fuh Yong Wong[6], Jingmei Li[1,2,*]

[1]*Laboratory of Women's Health and Genetics, Genome Institute of Singapore, A*STAR Research Entities, Singapore;* [2]*Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore;* [3]*Saw Swee Hock School of Public Health, National University of Singapore, Singapore;* [4]*Division of Medical Oncology, National Cancer Centre Singapore, Singapore, Singapore;* [5]*Department of Surgery, University Surgical Cluster, National University Hospital, Singapore, Singapore;* [6]*Division of Radiation Oncology, National Cancer Centre Singapore, Singapore, Singapore*

ARTICLE INFO

Article history:
Received 31 January 2023
Received in revised form
8 June 2023
Accepted 9 June 2023
Available online 16 June 2023

Keywords:
Breast cancer
Family history
Loss-of-function variants
Polygenic risk scores
Screening

ABSTRACT

Purpose: The benefit of using individual risk prediction tools to identify high-risk individuals for breast cancer (BC) screening is uncertain, despite the personalized approach of risk-based screening.

Methods: We studied the overlap of predicted high-risk individuals among 246,142 women enrolled in the UK Biobank. Risk predictors assessed include the Gail model (Gail), BC family history (FH, binary), BC polygenic risk score (PRS), and presence of loss-of-function (LoF) variants in BC predisposition genes. Youden J-index was used to select optimal thresholds for defining high-risk.

Results: In total, 147,399 were considered at high risk for developing BC within the next 2 years by at least 1 of the 4 risk prediction tools examined ($Gail_{2\text{-year}} > 0.5\%$: 47%, $PRS_{2\text{-yea}}$ r > 0.7%: 30%, FH: 6%, and LoF: 1%); 92,851 (38%) were flagged by only 1 risk predictor. The overlap between individuals flagged as high-risk because of genetic (PRS) and Gail model risk factors was 30%. The best-performing combinatorial model comprises a union of high-risk women identified by PRS, FH, and, LoF ($AUC_{2\text{-year}}$ [95% CI]: 62.2 [60.8 to 63.6]). Assigning individual weights to each risk prediction tool increased discriminatory ability.

Conclusion: Risk-based BC screening may require a multipronged approach that includes PRS, predisposition genes, FH, and other recognized risk factors.

The Article Publishing Charge (APC) for this article was paid by Genome Institute of Singapore.

*Correspondence and requests for materials should be addressed to Jingmei Li, Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01, Singapore 138672, Singapore. *Email address:* lijm1@gis.a-star.edu.sg

doi: https://doi.org/10.1016/j.gim.2023.100917

Introduction

Breast cancer (BC) accounts for 15.5% of the 4,429,323 cancer deaths observed globally among women.[1] Serial mammography screening efficiently identifies early-stage BC and reduces BC mortality.[2-6] Population-based mammography screening is usually targeted at women aged between 50 and 69 years.[7] However, more than half of the 2,261,419 BC cases diagnosed worldwide fall outside this age group (<50 years: 29.4%; ≥70 years: 22.4%) (GLOBOCAN, accessed September 19, 2022). In addition, inconsistent benefits-to-harm ratio limits the use of mammography screening in women outside the at-risk age group.

BC screening should ideally be performed when the risk of the disease is high enough to offset the harms of overdiagnosis and overtreatment.[2,8,9] Women below the age of 50 have a lower probability of developing BC than older women, but the forms of BC that do so are frequently more aggressive and have a worse prognosis.[10] Additionally, younger women are expected to live longer and have fewer comorbidities.[10] In contrast, older women (≥70 years) suffer from more comorbidities, casting doubts on the efficacy of mammography screening in reducing mortality. In addition, there are doubts regarding the efficacy of mammography screening among older women in reducing mortality because of a higher burden from noncancer comorbidities.[11-13] Furthermore, the risk of overdiagnosis and unnecessary treatment may compromise their quality of life and physical function.[14] "For whom does screening benefit" thus becomes an important question. The risk-benefit ratio for screening mammography for women outside the current target risk group may be tilted by personalized risk assessments and lead to better patient outcomes.[10]

The most efficient method to implement risk-based screening for BC is still being investigated:

- A higher risk of BC exists in those who have a family history (FH) of the disease that is likely brought on by genetic factors common lifestyle variables or other shared family traits.[15-17]
- The Gail model, which incorporates classic BC risk factors including age, age at the first occurrence of menstruation, age at first child, number of breast biopsies, history of atypical hyperplasia, and number of immediate family members with BC, is widely used and validated in many populations of different ancestry.[18-23] However, it should be noted that many other BC prediction models, such as BOADICEA, Tyrer-Cuzick, and BRCAPRO have been developed.[24,25]
- Studies have argued that mammographic density—ascertained from the appearance of the breast tissue on mammograms—is more strongly associated with BC risk than risk factors in the Gail model.[26-29]
- Individual risk of developing BC is caused by a particular genetic vulnerability (such rare loss-of-function [LoF] variants in *BRCA1* or *BRCA2*).[30-34] Large consortia efforts have identified other clinically useful BC predisposition genes (*ATM*, *CHEK2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, and *TP53*).[35]
- Polygenic risk scores (PRS) that sum up the effects of multiple common variants associated with BC have been implemented in pilot precision health initiatives that stratify individuals by their disease risk.[36-45]

With the continuous development and refinement of BC risk assessment tools, "How much does genetics add?" becomes of interest. Combining genetic data with standard risk instruments meaningfully enhances risk stratification and improves discriminatory value in mammography screening programs.[46,47] The collective role that various BC risk predictors play in disease prediction has also been previously studied. Hassanin et al examined FH, PRS, and LoF variants in BC predisposition genes, and concluded that they jointly improved risk stratification for breast and prostate cancers.[48] However, the extent of overlap between the high-risk individuals identified by different tools is unclear. Here, we assessed the proportion of women identified as at high risk of developing BC as ascertained by different BC risk assessment tools, in 246,142 women from the UK Biobank data set. In addition, we evaluated the proportion of high-risk individuals who eventually developed the malignancy.

Materials and Methods

Study population

The UK Biobank is a publicly available scientific database and research tool with comprehensive genetic and health data from about 500,000 individuals in the United Kingdom.[49] Participants were recruited between 2006 and 2010 via mail invitation (5% response rate). Aged between 40 and 70 years, enrolled with a general practitioner, and residing within 20 miles of 1 of 22 evaluation centers in England, Scotland, and Wales were requirements for participants. Our cohort was restricted to 264,741 female participants (application 86846) (Supplemental Figure 1).

BC polygenic risk score (PRS)

Genome-wide genetic data are available for 487,201 UK Biobank participants (Data-Field 22828). The UK Biobank genotyping project, quality control, imputation, and related processes have been previously described. (Supplemental Material) (Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*. 2017:166298. https://doi.org/10.1101/166298). The imputed genotypes are aligned to the + strand of the reference and the positions are in GRCh37 coordinates. The list of the 313 SNPs and associated weights included in the BC PRS is given in Supplemental Table 1.[36]

LoF variants in 9 BC risk genes

The targeted sequencing analyses using population-level exome OQFE variants (PLINK format - final release, ukb23158_c*_b0_v1) were conducted on the Research Analysis Platform (https://ukbiobank.dnanexus.com) (Supplemental Material).[50] Poor-quality variants were excluded. The quality control criteria require that at least 90% of all genotypes for a given variant—independent of variant allele zygosity—have a read depth of at least 10 (ie, depth ≥ 10). Missense and synonymous variants (annotated as missense (0/5), missense (5/5), missense (≥1/5), or synonymous) were excluded. LoF variants (n = 795) in 9 BC risk genes (*ATM*, n = 222; *BRCA1*, n = 115; *BRCA2*, n = 179; *CHEK2*, n = 60; *PALB2*, n = 85; *BARD1*, n = 49; *RAD51C*, n = 23; *RAD51D*, n = 47; or *TP53*, n = 15) were extracted.[35] The maximum minor allele frequency for LoF variants within the UK Biobank data set was set at 0.01. The resulting exome sequencing data set included 508 LoF variants in 254,635 females.

Non-genetic risk factors

Demographic and reproductive risk factors were obtained from the first instance: age at recruitment (Data-Field: 21022), race (Data-Field: 21000), age at menarche (Data-Field: 2714), parity (Data-Field: 2734), age at first childbirth (Data-Fields: 2754 and 3872), FH of BC (Data-Fields: 20110 for mother and 20111 for siblings), and menopausal status (Data-Field: 2724). FH of BC takes values 0 (no FH), 1 (where the individual's mother (Data-Fields: 20110) or at least 1 sibling (Data-Fields: 20111) had BC), or 2 (mother and at least 1 sibling had BC). Information on ever BC screening (Data-Field: 2674) was also retrieved.

BC case ascertainment

Invasive BC was determined using the 9th (174*) and 10th (C50*) versions of the International Classification of Diseases (Data-Fields 40013 "174 Malignant neoplasm of female breast" and 40006 "Malignant neoplasm of breast," respectively). Age at cancer diagnosis (Data-Fields: 20007 and 40008) was used to determine if the cancer diagnosis was before the age of 80. A total of 26 BCs were diagnosed at age 80 and above. Our focus is on risk-based BC screening for earlier diagnosis and treatment, the late age at diagnosis may be a case for overdiagnosis where the harms of treatment may outweigh the benefits; hence, these cases were considered noncase in our analysis. In situ BC cases were included as noncases.

A total of 253,953 females had information on both PRS and LoF variants (Supplemental Figure 1). The UK Biobank performs regular linkages to national cancer registries as a passive follow-up to capture cancer diagnosis. Censoring dates (dates estimated by the UK Biobank that information received from data providers is mostly complete) for National Health Services Digital and National Records of Scotland, National Health Services Central Register are 29 February 2020, and 31 January 2021, respectively. Prevalent cases refer to cancer diagnoses made before enrollment in the UK Biobank study (before the baseline assessment). Only the prevalent cancer is considered if a person was diagnosed with it both before and after baseline. If there were multiple diagnosis dates for a certain malignancy, the earliest date is used. If a participant was diagnosed with multiple cancers at the earliest date, each cancer type is separately counted. Because this analysis is aimed at predicting the risk of BC development among women without a previous BC diagnosis, women whose age at recruitment (Data-Fields: 21022) were older than their age at cancer diagnosis (ie, prevalent cases) (n = 7811) were excluded. The resulting analytical cohort consisted of 246,142 females, with 7620 incident cases of BC (the latest case was diagnosed in 2020).

Statistical analysis

Associations between risk factors of interest and invasive BC diagnosis before age 80 years were tested using χ^2 test for categorical variables and Kruskal-Wallis test for continuous variables.

Our outcomes of interest are invasive BC diagnosis within 2, 5, and 10 years post-study entry. Two-year absolute risks were examined because it was opined previously that "a risk model for effective screening should be designed to assess the risk of BC in the interval between the just-performed screen and the next scheduled screen to identify women who need supplemental screening."[51] For each period of interest, the corresponding x-year absolute risk (2-, 5-, and 10-year) was computed.

The individual's x-year absolute risk (2-year, 5-year, and 10-year) was predicted using the package (*BCRA* in R) for the Gail model.[18,52] The method to obtain the x-year absolute risk computed from PRS is described previously.[37] BC incidence rates from 2011 to 2015, and mortality rates of 2016, were used in the PRS absolute risk calculation.[53,54] The distributions of x-year absolute risks predicted by PRS and the Gail model are illustrated in Supplemental Figure 2.

To our knowledge, there is no consensus on the threshold to determine high-risk women based on the x-year absolute risks computed from PRS. We selected the threshold based on the highest Youden J-Index (*pROC* package in R).[55,56] This threshold optimization was repeated for the x-year absolute risks computed from the Gail model. If a woman was censored because of death before the end of the prediction period (ie, x-year) and before getting BC, she was considered a noncase in the logistic regression. Among the noncases, 403 (0.1% of 244,933 noncases) died within 2 years from recruitment, 2164 (0.9% of 242,882) died within 5 years, and 7001 (2.9% of 239,516) died within 10 years. FH of BC was treated as a binary variable (FH, yes or no)

when used as a risk predictor on its own; in the Gail model, values 0, 1, or 2+ were adopted. Women with predicted LoF pathogenic variants in at least 1 of the 9 BC predisposition genes (*ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *RAD51D*, *RAD51C*, and *TP53*) were considered high risk.

The women identified to be at high risk of developing BC according to the 4 risk prediction tools (PRS, Gail, FH, and LoF) were represented in a Venn diagram to illustrate the extent of overlap of at-risk individuals. The discriminatory ability of different risk prediction tools was assessed as single predictors and in combinations (union of high-risk women identified) by computing area under the receiver operating curve (AUC). The backward stepwise logistic regression analysis (R function step()) was used to examine the relations (beta coefficients) of the 4 risk prediction tools on BC risk (ie, the assigned weights of each prediction tool was the beta coefficients of the logistic model) (Supplemental Material).

To understand the absolute risk of developing BC in the different risk groups (low or high), we conducted time-to-event analyses using Cox proportional hazards models. Follow-up time was defined as the time between recruitment and the year of BC diagnosis, death, or the latest linkage year (2020), whichever comes earlier. Women were identified as at high risk based on the full model (PRS, Gail, FH, or LoF) and the model excluding the Gail model (PRS, FH, and LoF). The threshold selected was based on the highest Youden J-Index in the logistic regression model. Kaplan-Meier curves were plotted to observe the change in the proportion of women developing BC in the different risk groups.

R version 4.0.3 was used in all analyses.

Results

Study population characteristics

We studied 246,142 women without BC at study entry. The median age was 56 years (interquartile range: 50 to 63) (Table 1). The majority (n = 232,118, 94%) were of White background, were menopausal at study entry (n = 147,683, 60%), and attended BC screening (n = 194,441, 79%). As of 2020, 3% of the study population (n = 7620) developed invasive BC. The median age at diagnosis was 63 years (interquartile range: 57 to 69). Differences in the distribution of risk factors (age at menarche, age at first live birth, and number of children) by case status were small but statistically significant because of the large sample size (Table 1).

Optimal thresholds for the definition of high risk

The most optimal Youden J-Index was achieved with absolute risk cutoffs of 0.7% ($PRS_{2\text{-year}}$ AUC [95% CI]: 65.1 [63.6 to 66.7]) and 0.5% ($Gail_{2\text{-year}}$ AUC [95%CI]: 59.3 [57.7 to 60.9]). Supplemental Figure 3 summarizes the corresponding AUC values when considering 2-, 5-, and 10-years. The most optimal cutoffs were used for subsequent analyses.

Overlap of high-risk individuals identified by different BC risk assessment tools

The proportion of women flagged as high risk by $Gail_{2\text{-year}>0.5\%}$, $PRS_{2\text{-year}>0.7\%}$, FH, and LoF were 47% ($n$ = 115,986), 30% (n = 73,775), 6% (n = 15,770), and 1% (n = 3005), respectively (Figure 1). Supplemental Figure 4 shows corresponding Venn diagrams for 5-year and 10-year. Thirty-eight percent of the 246,142 women in the study were considered to be at high-risk by only 1 risk prediction tool ($PRS_{2\text{-year}>0.7\%}$ unique individuals = 28,630, $Gail_{2\text{-year}>0.5\%}$ unique individuals = 61,911, FH unique individuals = 1142, and LoF unique individuals = 1168).

Seventy-nine percent of the 1209 BC cases that developed within 2 years were identified to be at high risk by at least 1 of the 4 BC predictors examined. Using a 2-year absolute risk cutoff of 0.5% and 0.7%, the Gail model and PRS identified 60% (n = 728) and 52% (n = 632) of the cases, respectively. FH made up 12% (n = 145) of the BC cases that developed within the next 2 years of assessment; women with LoF variants made up less than 4% (n = 44). Although LoF variants are rare in the population ($n_{\text{high-risk, 10-year projection}}$ = 3005), more individuals develop BCs (n = 192, 6%) compared with the other 3 risk predictors (5014 BCs in 148,291 high-risk individuals, 3%) (Figure 2 and Supplemental Figure 4).

Improvement in the number of BCs identified within a high-risk group versus a random sample

When considered as single risk predictors, BC PRS was associated with the highest gain in the proportion of BC cases detected in the assessed period compared with the null line, followed by the Gail model (GAIL), first-degree FH of BC, and presence of LoF variants in high-penetrant BC genes (Figure 2). The best-performing combinatorial model comprises PRS, FH, and LoF ($AUC_{2\text{-year}}$ [95% CI]: 62.2 [60.8 to 63.6]) (Figure 2, Table 2).

Assigning weights to each risk prediction tool improves the discriminatory ability

To account for the effect overlap between the different tools, we examined whether the discriminatory ability changes when risk prediction tools were assigned different weights (Table 3). The best-performing backward stepwise logistic regression model retained all 4 risk prediction tools (Youden J-Index = 24.7, $AUC_{2\text{-year}}$ [95% CI]: 66.4 [64.8 to 67.9]). After manually removing the Gail model, the AUC corresponding to a model based on only PRS, LoF, and FH was found to be similar (66.3 [64.7 to 67.8]).

The median follow-up time for 246,142 women was 12 years (interquartile range: 11 to 13 years). Of the 7646

Table 1 Characteristics of study participants

	Breast Cancer Diagnosis Before Age 80 Years			
	All	No	Incident	
Characteristics	*N* = 246,142	*n* = 238,522	*n* = 7620	*P*
Median age at recruitment (IQR)	56 (50 to 63)	56 (50 to 63)	57 (51 to 63)	<.001
Median age at breast cancer diagnosis (IQR)			63 (57 to 69)	
Race				
White (includes British, Irish, and other White background)	232,118	224,843 (94%)	7275 (95%)	<.001
African American (includes African, Caribbean, Black or Black British, and other Black background)	4092	4017 (2%)	75 (1%)	
Chinese American (includes Chinese)	887	868 (0%)	19 (0%)	
Other Asian (includes Bangladeshi, Indian, Pakistani, Asian or Asian British, and other Asian background)	4092	3971 (2%)	121 (2%)	
Other (includes mixed, White and Black, White and Black African, White and Asian, unknown, and prefer not to answer)	4953	4823 (2%)	130 (2%)	
First degree family history of breast cancer (mother and siblings)				
None	230,372	223,519 (94%)	6853 (90%)	<.001
1	15,304	14,576 (6%)	728 (10%)	
2 or more	466	427 (0%)	39 (1%)	
Age at menarche, years				
≥14	87,577	84,981 (36%)	2596 (34%)	.009
12 to 13	103,892	100,562 (42%)	3330 (44%)	
≤11	47,014	45,534 (19%)	1480 (19%)	
Unknown	7659	7445 (3%)	214 (3%)	
Number of children				
0	45,798	44,291 (19%)	1507 (20%)	.002
1	32,833	31,765 (13%)	1068 (14%)	
2	107,513	104,218 (44%)	3295 (43%)	
3+	59,442	57,707 (24%)	1735 (23%)	
Unknown	556	541 (0%)	15 (0%)	
Age at first live birth, years				
No child	45,798	44,291 (19%)	1507 (20%)	<.001
≤19	21,460	20,905 (9%)	555 (7%)	
20 to 24	63,671	61,831 (26%)	1840 (24%)	
25 to 29	70,749	68,525 (29%)	2224 (29%)	
≥30	43,513	42,044 (18%)	1469 (19%)	
Unknown	951	926 (0%)	25 (0%)	
Menopausal status				
Yes	147,683	142,940 (60%)	4743 (62%)	<.001
No	59,266	57,586 (24%)	1680 (22%)	
Unknown	39,193	37,996 (16%)	1197 (16%)	
Ever attended breast cancer screening				
Yes	194,441	188,111 (79%)	6330 (83%)	<.001
No	50,979	49,707 (21%)	1272 (17%)	
Unknown	722	704 (0%)	18 (0%)	
Median polygenic risk score (IQR)	−0.315 (−0.730 to 0.096)	−0.324 (−0.737 to 0.087)	−0.055 (−0.473 to 0.365)	<.001
High-penetrance breast cancer genes (*ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *RAD51D*, *RAD51C*, or *TP53*)				
No	243,137	235,728 (99%)	7409 (97%)	<.001
Yes (at least 1 loss-of-function variant)	3005	2794 (1%)	211 (3%)	

Column percentages are shown within brackets.

IQR, interquartile range; *P*, *P* value from χ^2 test (categorical variable) or Kruskal-Wallis test (continuous variable).

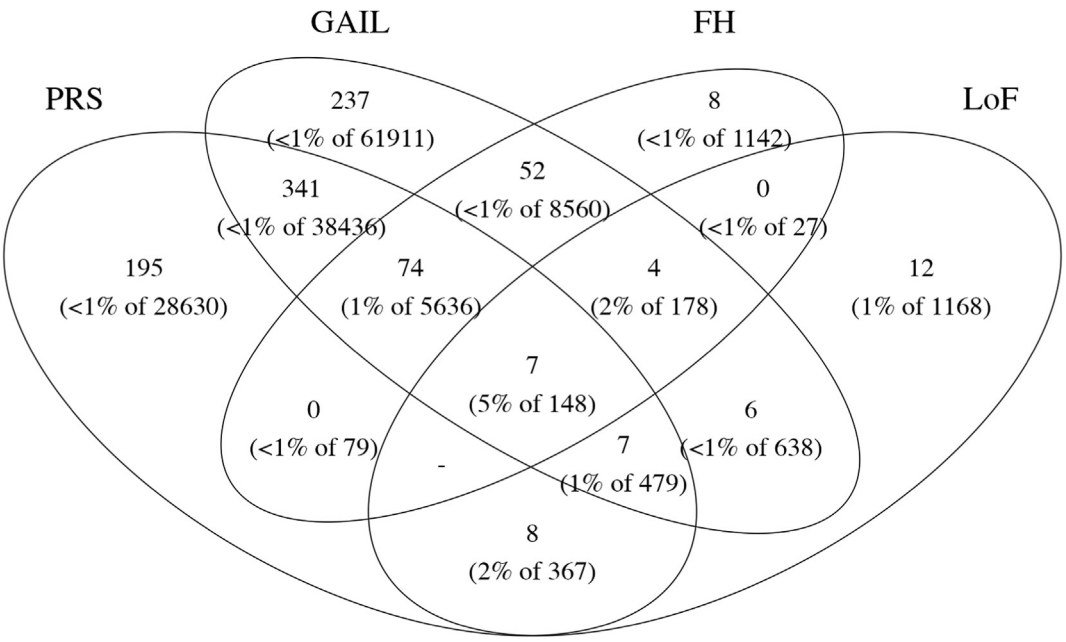

Figure 1 Number of incident invasive breast cancer events by the 4 risk prediction tools represented in a Venn diagram. The percentages of breast cancer events in high-risk women are shown within brackets. High-risk women were identified using the following criteria: (1) *x*-year absolute risk above threshold as predicted by polygenic risk score (PRS: >0.7% for 2-year absolute risk, respectively), (2) *x*-year absolute risk above threshold as predicted by the Gail model (GAIL: >0.5% for 2-year absolute risk, respectively), (3) family history of breast cancer (yes, FH), and (4) presence of at least one loss-of-function variants (LoF) in any of the 9 breast cancer predisposition genes (*ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *RAD51D*, *RAD51C*, and/or *TP53*). GAIL, the Gail model; FH, family history of breast cancer; LoF, at least one loss-of-function variant; PRS, polygenic risk score.

incident cases of BC (median follow-up time: 5 years, interquartile range: 3 to 8 years), 474 were diagnosed in the same year of recruitment. The Kaplan-Meier curves showed significant separation between the low- and the high-risk groups when we based the logistic regression model on the 2-, 5- and 10-year absolute risk of PRS and the Gail model (Supplemental Figure 5). At 10 years of follow-up, 98% of women identified as low risk and 96% of the high-risk group were disease-free. High-risk women were 2.2 times (95% CI: 2.1 to 2.3) as likely to develop BC as those in the low-risk group (Supplemental Table 2).

Discussion

Stratifying population-level service users by their individual BC risk may improve resource utilization and alleviate the issue of overdiagnosis.[57] We identified individuals at high risk of developing BC based on different established genetic (PRS and LoF) and non-genetic (FH and the Gail model) risk calculators in the large UK Biobank study. Two-, five-, ten-year, and lifetime BC absolute risks were computed. The analysis associated with 2-year BC absolute risk was associated with the highest discriminatory value. Among the 246,142 women in the analytical cohort, 147,399 were considered at high risk for developing BC within the next 2 years by at least 1 of the 4 BC risk assessment tools examined. Among the high-risk individuals, 92,851 (38%) were flagged by only 1 risk predictor. Seventy-nine percent of the BCs that did develop within the next 2 years were from the high-risk group. The union of high-risk individuals identified by PRS, FH, and LoF yielded the best improvement in the number of BC cases detected when compared with a random sample. Assigning individual weights to each risk prediction tool appeared to increase the discriminatory ability.

Our observation that a large proportion of women are uniquely flagged as high-risk by only 1 risk assessment tool suggests that BC screening may benefit from including genetic and non-genetic risk factors. BC PRS exerts an effect distinct from traditional risk factors.[58,59] The effect of PRS on BC risk is known not to be strongly correlated with FH[36,39,60-62] and largely independent of other known risk factors for BC, such as mammographic density,[63] lifestyle factors,[59,60,64,65] reproductive factors, and hormone use.[60,65]

Considering different risk prediction models in tandem improves performance.[46,47,66] In a study comprising 126,894 women, the joint predictive model performed better than PRS or non-genetic risk score alone.[46] In the FinnGen study, PRS improves risk prediction in women with a FH of BC.[66] In the Predicting the Risk of Cancer At Screening study (PROCAS study), PRS improved risk stratification significantly when compared with a model based only on mammographic density and conventional risk factors; however, the inclusion of gene panels showed no appreciable effect.[47] In our study, the best discriminatory ability

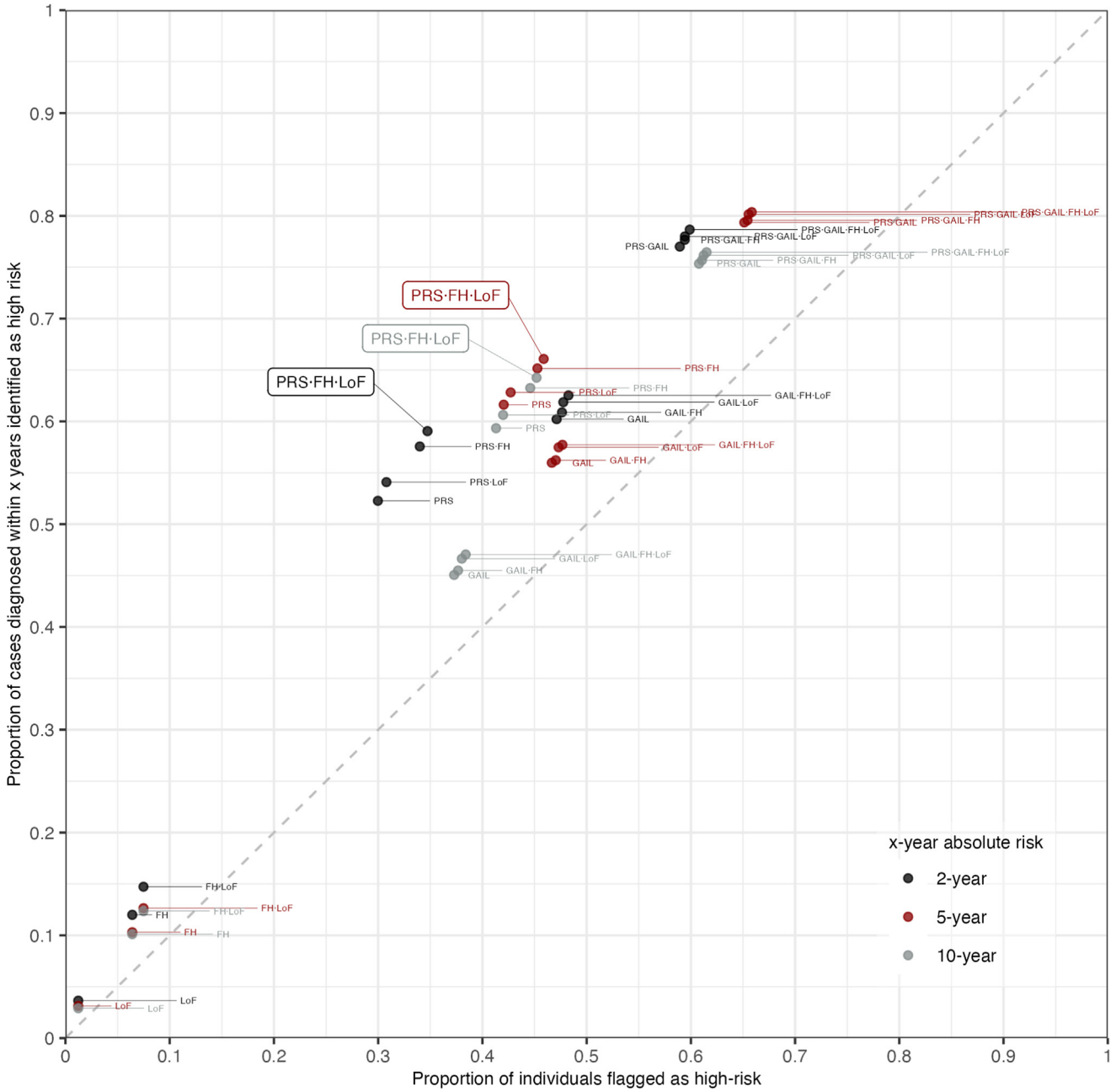

Figure 2 Comparison of how different combinations of breast cancer risk assessment tools perform in the UK Biobank (n = 246,142 females, median age [IQR] = 56 [50 to 63] years). The figure shows the proportion of individuals flagged as high-risk by different breast cancer risk assessment tools (x-axis) and the proportion of cases diagnosed within x years identified as high risk (where x is 2, 5, or 10; y-axis). Breast cancer polygenic risk score (PRS) was associated with the highest gain in the proportion of breast cancer cases detected in the assessed period compared to the null line, followed by the Gail model (GAIL), first-degree family history of breast cancer (FH), and presence of at least one loss-of-function variants in high-penetrant breast cancer genes (LoF). The best-performing combinatorial model (in boxed labels) comprises PRS, FH, and LoF. FH, family history of breast cancer; LoF, at least one loss-of-function variant; PRS, polygenic risk score.

was associated with a combinatorial model including PRS, FH, and LoF.

We show that risk stratification using PRS, FH, and LoF offered the best gain in terms of the number of BC cases detected in a high-risk population when compared with a random sample. Fifteen percent (63% in high-risk women vs 48% of random selection) more cases may be identified by looking within the high-risk women (48% of the population) compared with taking a random sample of the same proportion. This result agrees with previous works examining the impact of using multiple risk prediction tools in risk-based BC screening scenarios.[67-69] Notably, Darabi et al estimated that a customized screening strategy with input from multiple risk models (eg, conventional risk factors,

Table 2 Discriminatory ability and performance measures when women were flagged as high-risk by taking the union of the risk predictor combination selected in Figure 2

						95% CI of AUC					
x-Year Absolute Risk	PRS[a]	Model	Sensitivity	Specificity	AUC	Lower	Upper	FPR	FNR	TPR	TNR
2-year	0.7	PRS ⋃ FH ⋃ LoF	59.1	65.4	62.2	60.8	63.6	99.2	0.3	0.8	99.7
5-year	1.4	PRS ⋃ FH ⋃ LoF	66.1	54.4	60.2	59.4	61.0	98.1	0.8	1.9	99.2
10-year	2.9	PRS ⋃ FH ⋃ LoF	64.2	55.3	59.8	59.2	60.4	96.2	1.8	3.8	98.2

AUC, area under the receiver operating curve; *FH*, family history of breast cancer; *FNR*, false negative rate; *FPR*, false positive rate; *GAIL*, the Gail model; *LoF*, at least one loss-of-function variant; *PRS*, polygenic risk score; *TNR*, true negative rate; *TPR*, true positive rate.

[a]*x*-Year absolute risk threshold to define high risk.

mammographic density, and PRSs) captures 10% more cases than an age-based approach.[67] In a scenario where BC screening in the UK is adapted to screen women aged 35 to 79 years based on PRS rather than age alone, it is anticipated that the proportion of women eligible for screening may be decreased by 24%, resulting in a 14% reduction in screen-detectable cases.[70]

An issue that emerged from this analysis is the large number of women considered to be at high-risk by PRS and the Gail model. All BC cases would be detected if all women were considered to be at high-risk, which defeats the aim of risk stratification. Hence, though our results show that PRS and Gail are able to identify BC cases missed by other risk predictors, it comes at a larger at-risk pool of women who may or may not develop BCs in the future.

In previous literature, annual BC incidence rates can be 2 to 10 times higher among women considered high risk than the general population.[71] High risk for BC has been defined as having a greater than or equal to 20% lifetime risk by the American Cancer Society guidelines, or a 10-year risk of ≥8% according to the National Institute of Health and Care Excellence risk assessment tool.[72,73] Studies have shown that only a small percentage of women are at high risk for developing BC because of various measures. Apart from groups with frequent founder pathogenic variants, such as the Ashkenazi Jewish community, the population frequency of pathogenic *BRCA1/2* variants has been reported to be 1:400.[74] Approximately 1 in 10 women in the Breast Cancer Surveillance Consortium study reported first-degree FH of BC.[75] As reported by Niell et al, 13.7% of the screening population studied were at elevated risk (>20% absolute lifetime risk) according to 1 of 3 risk prediction models (ie, modified Gail, Tyrer-Cuzick version 7, and BRCAPRO) studied.[73] In another report, Evans et al stated that only about 1% of women in the population meet the predetermined definition of 8% 10-year risk or 30% lifetime risk based solely on FH and conventional risk factors.[72] However, this can increase to 6% when combined with mammographic density and common BC genetic risk variants.[72] It is important to note that the proportion of women classified under high-risk varies depending on the study and the criteria used to define high-risk. Because the UK Biobank cohort is not a high-risk cohort, cost-effectiveness will be a consideration when choosing a risk threshold.

To take different effect sizes into account, we assigned weights to individual risk assessment tools and found that the discriminatory value improved (95% CI of the AUCs did not overlap). Nonetheless, risk thresholds specific to this data set were derived by optimizing a statistical criterion (eg, the Youden index) for the purpose of evaluating the extent of overlap in high-risk individuals flagged by different prediction models. The risk definitions in this study may be different from widely accepted standards in clinical practice. For example, according to the Gail model, a 5-year absolute risk of 1.67% or higher is the U.S. Food and Drug Administration guideline for taking a risk-lowering drug to reduce BC risk. The threshold for what is considered at high risk will ultimately depend on the intended interventions (eg, chemoprevention, more frequent screenings, or targeted invitation to attend screenings) and health care resources available in each country. It should also be noted that dichotomized risk stratification implies a loss of information—everyone at high risk may be construed and regarded as though they have the same risk.[76] However, in the course of implementation, a calibrated continuous risk provides for more nuanced decision making at the individual level.

It has been shown that BC PRS have largely similar predictive performance in Asian and European women.[37,46] However, calibration performance is important to consider for risk prediction tools for the translational implementation of PRS.[77] Evans et al examined 2 PRS developed and validated in populations of European ancestry in underrepresented populations and reported that both PRS overstated BC risk across all racial and ethnic groups (~40% overestimation in combined populations of non-European ancestry).[78] Similar observations regarding the overestimation of risks, especially at the higher risk extremes, have been made in studies of Asian populations.[79,80] Regardless, further studies are required to examine the association of BC variants in non-European populations and to refine or redefine ethnic-specific scores.[81]

The UK Biobank sample size offers significant statistical power, well-documented and defined data collection processes, and case identification by linking to national cancer registries. As with any cohort study, the potential for selection bias, such as a healthy volunteer selection bias, cannot be dismissed.[82] Participants in the UK Biobank are known to be of higher economic status and have fewer

Table 3 Discriminatory ability and performance measures of different strategies in identifying high-risk women

	Absolute Risk Cutoff (%)		Model Statistics											
x-Year Absolute Risk	PRS	GAIL	Model	Threshold (*P*)	Sensitivity	Specificity	Youden J-statistics	AUC	95% CI of AUC Lower	95% CI of AUC Upper	FPR	FNR	TPR	TNR
Full model (also the best by stepwise backward selection)														
2-year	0.7	0.5	Logit(p): −6.29 + 0.70 PRS + 0.80 GAIL + 0.16 FH + 1.03 LoF	.5	61.4	63.4	24.7	66.4	64.8	67.9	99.2	0.3	0.8	99.7
5-year	1.4	1.4	Logit(p): −5.15 + 0.29 PRS + 0.21 GAIL + 0.15 FH + 0.90 LoF	1.2	59.6	61.5	21.1	64.4	63.5	65.4	98.0	0.9	2.0	99.1
10-year	2.9	3.0	Logit(p): −4.43 + 0.15 PRS + 0.10 GAIL + 0.16 FH + 0.85 LoF	2.5	59.0	60.6	19.6	63.5	62.9	64.2	96.0	1.8	4.0	98.2
Model without GAIL														
2-year	0.7	-	Logit(p): −5.91 + 0.75 PRS + 0.59 FH + 1.02 LoF	0.5	56.2	68.3	24.5	66.3	64.7	67.8	99.1	0.3	0.9	99.7
5-year	1.4	-	Logit(p): −4.88 + 0.30 PRS + 0.43 FH + 0.90 LoF	1.3	53.6	67.2	20.7	64.4	63.5	65.4	97.9	0.9	2.1	99.1
10-year	2.9	-	Logit(p): −4.16 + 0.15 PRS + 0.43 FH + 0.85 LoF	2.5	55.8	63.8	19.6	63.5	62.9	64.2	95.9	1.9	4.1	98.1

The models show beta weights from a stepwise logistic regression with backward removal predicting breast cancer risk using the 4 different risk prediction tools.

AUC, area under the receiver operating curve; *FH*, family history of breast cancer; *FNR*, false negative rate; *FPR*, false positive rate; *GAIL*, the Gail model; *LoF*, at least one loss-of-function variant; *P*, probability from the combinatorial model; *PRS*, polygenic risk scores; *TNR*, true negative rate; *TPR*, true positive rate.

lifestyle risk factors.[83] In addition, UK Biobank participants may be more health-conscious, have fewer comorbidities, and, among older women, be associated with lower all-cause death rates than the general population.[82] In the selection of risk threshold using the Youden J-index, individuals who died within the prediction period were treated as noncases. However, because the number of censored cases was small, it is unlikely to have much of an impact on the results. The generalizability of our findings may be limited to women of European heritage. Limited access to non-genetic risk factors in the UK Biobank, such as detailed FH, number of breast biopsies, and history of atypical hyperplasia may explain the poor performance of the Gail model in BC risk stratification. Mammographic density, a strong risk factor for BC, is also not available as a variable for BC risk assessment. As information on BC stage and hormone-receptor subtype are not available in the UK Biobank, we were unable to subset the analyses by tumor features.

Our findings suggest that risk-based BC screening programs may benefit from a multipronged approach that includes PRS, pathogenic variants in BC predisposition genes, FH, and other recognized risk factors. In this regard, comprehensive multipronged models are already being used in clinical practice, especially in high-risk populations.[84,85] However, a potential downside of these models is that their inputs are more time-consuming to collect than the Gail model. Nonetheless, to be successful, screening programs require significant health resources, a strong infrastructure, and capability within the country's health care system.[86] There are other remaining issues regarding optimal risk thresholds, how participants are informed of risk assessment findings, and how future policies may be shaped before the potential of precision screening for BC is realized.

Data Availability

Publicly available data from the UK Biobank study (application 86846) were analyzed in this study. The data sets are available to researchers through an open application via https://www.ukbiobank.ac.uk/register-apply/.

Acknowledgments

This work was supported by the A*STAR Computational Resource Centre through the use of its high-performance computing facilities.

Funding

This work is supported by the Agency for Science, Technology and Research (A*STAR) and the PRECISION Health Research, Singapore Clinical Implementation Pilot (PRECISE BREATHE CIP) Fund.

Author Information

Conceptualization: J.L., P.J.H., E.H.L., M.H., F.Y.W.; Data curation: J.L.; Formal analysis: P.J.H., J.L.; Funding acquisition: J.L.; Methodology: J.L., P.J.H.; Resources: J.L.; Visualization: J.L., P.J.H.; Writing-original draft: J.L., P.J.H.; Writing-review and editing: J.L., P.J.H., E.H.L., M.H., F.Y.W.

Ethics Declaration

UK Biobank has approval from the National Health Services National Research Ethics Service (16/NW/0274) and the North West Multi-Centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval (11/NW/0382). Each participant provided written informed consent. Permission to access and analyze UK Biobank data (Project Application Number 86846) was approved by the UK Biobank according to their established access procedures. Our research project also received approval from the A*STAR Institutional Review Board (2022-041).

Conflict of Interest

The authors declare no conflicts of interest.

Additional Information

The online version of this article (https://doi.org/10.1016/j.gim.2023.100917) contains supplemental material, which is available to authorized users.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. http://doi.org/10.3322/caac.21660
2. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med*. 2012;367(21):1998-2005. http://doi.org/10.1056/NEJMoa1206809
3. de Koning HJ. Mammographic screening: evidence from randomised controlled trials. *Ann Oncol*. 2003;14(8):1185-1189. http://doi.org/10.1093/annonc/mdg319
4. Mandrik O, Zielonke N, Meheus F, et al. Systematic reviews as a "lens of evidence": determinants of benefits and harms of breast cancer screening. *Int J Cancer*. 2019;145(4):994-1006. http://doi.org/10.1002/ijc.32211
5. Coleman C. Early detection and screening for breast cancer. *Semin Oncol Nurs*. 2017;33(2):141-155. http://doi.org/10.1016/j.soncn.2017.02.009
6. World Health Organization. *WHO position paper on mammography screening*. World Health Organization; 2014. ISBN: 978 92 4 150793 6 Accessed February 2, 2023. https://www.who.int/publications/i/item/who-position-paper-on-mammography-screening

7. Ebell MH, Thai TN, Royalty KJ. Cancer screening recommendations: an international comparison of high income countries. *Public Health Rev*. 2018;39:7. http://doi.org/10.1186/s40985-018-0080-0
8. Esserman LJ, Thompson IM, Reid B, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncol*. 2014;15(6):e234-e242. http://doi.org/10.1016/S1470-2045(13)70598-9
9. Kowalski AE. Mammograms and mortality: how has the evidence evolved? *J Econ Perspect*. 2021;35(2):119-140. http://doi.org/10.1257/jep.35.2.119
10. Grimm LJ, Avery CS, Hendrick E, Baker JA. Benefits and risks of mammography screening in women ages 40 to 49 years. *J Prim Care Community Health*. 2022;13:21501327211058322. http://doi.org/10.1177/21501327211058322
11. Richman IB, Gross CP. New breast cancer screening technologies in older women-is it time to pump the brakes? *JAMA Intern Med*. 2019;179(3):289-290. http://doi.org/10.1001/jamainternmed.2018.7767
12. Demb J, Abraham L, Miglioretti DL, et al. Cumulative incidence of non-breast cancer mortality and breast cancer risk by comorbidity and age among older women undergoing screening mammography: the Medicare-linked breast cancer surveillance consortium cohort study. *Cancer Epidemiol Biomarkers Prev*. 2018;27(3):358. http://doi.org/10.1158/1055-9965.EPI-18-0065, 358.
13. Brawley OW. On breast cancer screening in older women. *Ann Intern Med*. 2022;175(1):127-128. http://doi.org/10.7326/M21-4235
14. de Glas NA, de Craen AJ, Bastiaannet E, et al. Effect of implementation of the mass breast cancer screening programme in older women in the Netherlands: population based study. *BMJ*. 2014;349:g5410. http://doi.org/10.1136/bmj.g5410
15. Cole P. What should the physician ask? *Cancer*. 1976;37(1 suppl):434-436. http://doi.org/10.1002/1097-0142(197601)37:1+<434::aid-cncr2820370706>3.0.co;2-o
16. Haagensen CD. Family history of breast carcinoma in women predisposed to develop breast Carcinoma1. *J Natl Cancer Inst*. 1972;48(4):1025-1027.
17. Möller S, Mucci LA, Harris JR, et al. The heritability of breast cancer among women in the Nordic twin study of cancer. *Cancer Epidemiol Biomarkers Prev*. 2016;25(1):145-150. http://doi.org/10.1158/1055-9965.EPI-15-0913
18. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879-1886. http://doi.org/10.1093/jnci/81.24.1879
19. Bondy ML, Newman LA. Assessing breast cancer risk: evolution of the Gail Model. *J Natl Cancer Inst*. 2006;98(17):1172-1173. http://doi.org/10.1093/jnci/djj365
20. Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst*. 1999;91(18):1541-1548. http://doi.org/10.1093/jnci/91.18.1541
21. 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. *J Natl Cancer Inst*. 2001;93(5):358-366. http://doi.org/10.1093/jnci/93.5.358
22. Ho PJ, Wong FY, Chay WY, et al. Breast cancer risk stratification for mammographic screening: a nation-wide screening cohort of 24,431 women in Singapore. *Cancer Med*. 2021;10(22):8182-8191. http://doi.org/10.1002/cam4.4297
23. Bondy ML, Newman LA. Breast cancer risk assessment models: applicability to African-American women. *Cancer*. 2003;97(1suppl):230-235. http://doi.org/10.1002/cncr.11018
24. 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 Res Treat*. 2017;164(2):263-284. http://doi.org/10.1007/s10549-017-4247-z
25. Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. *J Natl Cancer Inst*. 2010;102(10):680-691. http://doi.org/10.1093/jnci/djq088
26. Boyd NF, Martin LJ, Rommens JM, et al. Mammographic density: a heritable risk factor for breast cancer. *Methods Mol Biol*. 2009;472:343-360. http://doi.org/10.1007/978-1-60327-492-0_15
27. Boyd NF, Rommens JM, Vogt K, et al. Mammographic breast density as an intermediate phenotype for breast cancer. *Lancet Oncol*. 2005;6(10):798-808. http://doi.org/10.1016/S1470-2045(05)70390-9
28. Pettersson A, Graff RE, Ursin G, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. *J Natl Cancer Inst*. 2014;106(5):dju078. http://doi.org/10.1093/jnci/dju078
29. Boyd NF, Martin LJ, Yaffe MJ, Minkin S. Mammographic density and breast cancer risk: current understanding and future prospects. *Breast Cancer Res*. 2011;13(6):223. http://doi.org/10.1186/bcr2942
30. Ang BH, Ho WK, Wijaya E, et al. Predicting the likelihood of carrying a BRCA1 or BRCA2 mutation in Asian patients with breast cancer. *J Clin Oncol*. 2022;40(14):1542-1551. http://doi.org/10.1200/JCO.21.01647
31. Domchek SM, Eisen A, Calzone K, Stopfer J, Blackwood A, Weber BL. Application of breast cancer risk prediction models in clinical practice. *J Clin Oncol*. 2003;21(4):593-601. http://doi.org/10.1200/JCO.2003.07.007
32. Barcenas CH, Hosain GM, Arun B, et al. Assessing BRCA carrier probabilities in extended families. *J Clin Oncol*. 2006;24(3):354-360. http://doi.org/10.1200/JCO.2005.02.2368
33. Lee A, Mavaddat N, Wilcox AN, et al. Boadicea: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med*. 2019;21(8):1708-1718. http://doi.org/10.1038/s41436-018-0406-9
34. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA*. 2017;317(23):2402-2416. http://doi.org/10.1001/jama.2017.7112
35. Breast cancer Association Consortium: Dorling L, Carvalho S, Allen J, et al. Breast cancer risk genes – Association Analysis in more than 113, 000 women. *N Engl J Med*. 2021;384(5):428-439. http://doi.org/10.1056/NEJMoa1913948
36. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*. 2019;104(1):21-34. http://doi.org/10.1016/j.ajhg.2018.11.002
37. Ho WK, Tan MM, Mavaddat N, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun*. 2020;11(1):3833. http://doi.org/10.1038/s41467-020-17680-w
38. Ho WK, Tai MC, Dennis J, et al. Polygenic risk scores for prediction of breast cancer risk in Asian populations. *Genet Med*. 2022;24(3):586-600. http://doi.org/10.1016/j.gim.2021.11.008
39. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5):djv036. http://doi.org/10.1093/jnci/djv036
40. Liu J, Ho PJ, Tan THL, et al. Breast screening tailored for her (BREATHE)-a study protocol on personalised risk-based breast cancer screening programme. *PLoS ONE*. 2022;17(3):e0265965. http://doi.org/10.1371/journal.pone.0265965
41. Roux A, Cholerton R, Sicsic J, et al. Study protocol comparing the ethical, psychological and socio-economic impact of personalised breast cancer screening to that of standard screening in the "My Personal Breast Screening" (MyPeBS) randomized clinical trial. *BMC Cancer*. 2022;22(1):507. http://doi.org/10.1186/s12885-022-09484-6
42. Brooks JD, Nabi HH, Andrulis IL, et al. Personalized risk assessment for prevention and early detection of breast cancer: integration and implementation (PERSPECTIVE I&I). *J Pers Med*. 2021;11(6):511. http://doi.org/10.3390/jpm11060511
43. Gabrielson M, Eriksson M, Hammarström M, et al. Cohort profile: the Karolinska mammography project for risk prediction of breast cancer (KARMA). *Int J Epidemiol*. 2017;46(6):1740-1741g. http://doi.org/10.1093/ije/dyw357
44. Esserman LJ, Study W, Athena I. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *npj Breast Cancer*. 2017;3:34. http://doi.org/10.1038/s41523-017-0035-5

45. Keane H, Huilgol YS, Shieh Y, et al. Development and pilot of an online, personalized risk assessment tool for a breast cancer precision medicine trial. *npj Breast Cancer*. 2021;7(1):78. http://doi.org/10.1038/s41523-021-00288-8
46. Yang Y, Tao R, Shu X, et al. Incorporating polygenic risk scores and nongenetic risk factors for breast cancer risk prediction among Asian women. *JAMA Netw Open*. 2022;5(3):e2149030. http://doi.org/10.1001/jamanetworkopen.2021.49030
47. Evans DGR, van Veen EM, Harkness EF, et al. Breast cancer risk stratification in women of screening age: incremental effects of adding mammographic density, polygenic risk, and a gene panel. *Genet Med*. 2022;24(7):1485-1494. http://doi.org/10.1016/j.gim.2022.03.009
48. Hassanin E, May P, Aldisi R, et al. Breast and prostate cancer risk: the interplay of polygenic risk, rare pathogenic germline variants, and family history. *Genet Med*. 2022;24(3):576-585. http://doi.org/10.1016/j.gim.2021.11.009
49. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. http://doi.org/10.1371/journal.pmed.1001779
50. Backman JD, Li AH, Marcketta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599(7886):628-634. http://doi.org/10.1038/s41586-021-04103-z
51. Eriksson M, Conant EF, Kontos D, Hall P. Risk assessment in population-based breast cancer screening. *J Clin Oncol*. 2022;40(20):2279-2280. http://doi.org/10.1200/JCO.21.02827
52. Banegas MP, John EM, Slattery ML, et al. Projecting individualized absolute invasive breast cancer risk in US Hispanic women. *J Natl Cancer Inst*. 2016;109(2):djw215. http://doi.org/10.1093/jnci/djw215
53. Cancer Incidence Rates for Combined Years by: Site(Third Digit), Sex and Age, England, 2011-15. Accessed February 2, 2023. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/adhocs/007386cancerincidenceratesforcombinedyearsbysitethirddigitsexandageengland2011to2015
54. Incidence of, and Mortality from Breast Cancer in England, 2004 to 2017. Accessed February 2, 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/adhocs/009444incidenceofandmortalityfrombreastcancerinengland2004to2017
55. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35. http://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3
56. Schisterman EF, Faraggi D, Reiser B, Hu J. Youden Index and the optimal threshold for markers with mass at zero. *Stat Med*. 2008;27(2):297-315. http://doi.org/10.1002/sim.2993
57. Gierach GL, Choudhury PP, García-Closas M. Toward risk-stratified breast cancer screening: considerations for changes in screening guidelines. *JAMA Oncol*. 2020;6(1):31-33. http://doi.org/10.1001/jamaoncol.2019.3820
58. Ho PJ, Ho WK, Khng AJ, et al. Overlap of high-risk individuals predicted by family history, and genetic and non-genetic breast cancer risk prediction models: implications for risk stratification. *BMC Med*. 2022;20(1):150. http://doi.org/10.1186/s12916-022-02334-z
59. Kapoor PM, Mavaddat N, Choudhury PP, et al. Combined associations of a polygenic risk score and classical risk factors with breast cancer risk. *J Natl Cancer Inst*. 2021;113(3):329-337. http://doi.org/10.1093/jnci/djaa056
60. Hou C, Xu B, Hao Y, Yang D, Song H, Li J. Development and validation of polygenic risk scores for prediction of breast cancer and breast cancer subtypes in Chinese women. *BMC Cancer*. 2022;22(1):374. http://doi.org/10.1186/s12885-022-09425-3
61. Li H, Feng B, Miron A, et al. Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. *Genet Med*. 2017;19(1):30-35. http://doi.org/10.1038/gim.2016.43
62. Wang Y, Zhu M, Ma H, Shen H. Polygenic risk scores: the future of cancer risk prediction, screening, and precision prevention. *Medical Review*. 2021;1(2):129-149. http://doi.org/10.1515/mr-2021-0025
63. 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. http://doi.org/10.1001/jamaoncol.2017.4881
64. Arthur RS, Wang T, Xue X, Kamensky V, Rohan TE. Genetic factors, adherence to healthy lifestyle behavior, and risk of invasive breast cancer among women in the UK Biobank. *J Natl Cancer Inst*. 2020;112(9):893-901. http://doi.org/10.1093/jnci/djz241
65. 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. *Int J Epidemiol*. 2018;47(2):526-536. http://doi.org/10.1093/ije/dyx242
66. Mars N, Widén E, Kerminen S, et al. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun*. 2020;11(1):6383. http://doi.org/10.1038/s41467-020-19966-5
67. Darabi H, Czene K, Zhao W, Liu J, Hall P, Humphreys K. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Res*. 2012;14(1):R25. http://doi.org/10.1186/bcr3110
68. Zhang X, Rice M, Tworoger SS, et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: a nested case-control study. *PLoS Med*. 2018;15(9):e1002644. http://doi.org/10.1371/journal.pmed.1002644
69. Shieh Y, Hu D, Ma L, et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res Treat*. 2016;159(3):513-525. http://doi.org/10.1007/s10549-016-3953-2
70. Pashayan N, Duffy SW, Chowdhury S, et al. Polygenic susceptibility to prostate and breast cancer: implications for personalized screening. *Br J Cancer*. 2011;104(10):1656-1663. http://doi.org/10.1038/bjc.2011.118
71. Vogel VG. High-risk populations as targets for breast cancer prevention trials. *Prev Med*. 1991;20(1):86-100. http://doi.org/10.1016/0091-7435(91)90009-s
72. Evans DG, Howell SJ, Howell A. Personalized prevention in high risk individuals: managing hormones and beyond. *Breast*. 2018;39:139-147. http://doi.org/10.1016/j.breast.2018.03.009
73. Niell BL, Augusto B, McIntyre M, et al. Distribution of estimated lifetime breast cancer risk among women undergoing screening mammography. *AJR Am J Roentgenol*. 2021;217(1):48-55. http://doi.org/10.2214/AJR.20.23333
74. Maxwell KN, Domchek SM, Nathanson KL, Robson ME. Population frequency of germline BRCA1/2 mutations. *J Clin Oncol*. 2016;34(34):4183-4185. http://doi.org/10.1200/JCO.2016.67.0554
75. Durham DD, Abraham LA, Roberts MC, et al. Breast cancer incidence among women with a family history of breast cancer by relative's age at diagnosis. *Cancer*. 2022;128(24):4232-4240. http://doi.org/10.1002/cncr.34365
76. Wynants L, van Smeden M, McLernon DJ, et al. Three myths about risk thresholds for prediction models. *BMC Med*. 2019;17(1):192. http://doi.org/10.1186/s12916-019-1425-3
77. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230. http://doi.org/10.1186/s12916-019-1466-7
78. Evans DG, van Veen EM, Byers H, et al. The importance of ethnicity: are breast cancer polygenic risk scores ready for women who are not of White European origin? *Int J Cancer*. 2022;150(1):73-79. http://doi.org/10.1002/ijc.33782
79. Jee YH, Ho WK, Park S, et al. Polygenic risk scores for prediction of breast cancer in Korean women. *Int J Epidemiol*. 2023;52(3):796-805. http://doi.org/10.1093/ije/dyac206
80. Ho PJ, Tan IB, Chong DQ, et al. Polygenic risk scores for the prediction of common cancers in East Asians: a population-based prospective cohort study. *eLife*. 2023;12:e82608. http://doi.org/10.7554/eLife.82608
81. Yanes T, Young MA, Meiser B, James PA. Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field. *Breast Cancer Res*. 2020;22(1):21. http://doi.org/10.1186/s13058-020-01260-3

82. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026-1034. http://doi.org/10.1093/aje/kwx246
83. Hanlon P, Jani BD, Nicholl B, Lewsey J, McAllister DA, Mair FS. Associations between multimorbidity and adverse health outcomes in UK Biobank and the SAIL Databank: a comparison of longitudinal cohort studies. *PLoS Med*. 2022;19(3):e1003931. http://doi.org/10.1371/journal.pmed.1003931
84. Brédart A, Kop JL, Antoniou AC, et al. Clinicians' use of breast cancer risk assessment tools according to their perceived importance of breast cancer risk factors: an international survey. *J Community Genet*. 2019;10(1):61-71. http://doi.org/10.1007/s12687-018-0362-8
85. Pal Choudhury P, Brook MN, Hurson AN, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. *Breast Cancer Res*. 2021;23(1):22. http://doi.org/10.1186/s13058-021-01399-7
86. Wedderburn S, McVeigh TP. Should all individuals be screened for genetic predisposition to cancer? *Genet Res (Camb)*. 2021;2021:6611963. http://doi.org/10.1155/2021/6611963