How much does genetics add to screening? Breast cancer risk stratification using genetic and non-genetic risk assessment tools for 246,142 women in the UK Biobank.

Jingmei Li (lijm1@gis.a-star.edu.sg)
Genome Institute of Singapore
Peh Joo Ho
Genome Institute of Singapore https://orcid.org/0000-0002-3017-4023
Elaine Lim
National Cancer Centre Singapore
Mikael Hartman
National University of Singapore
Fuh Yong Wong
National Cancer Centre Singapore

Article

Keywords:

Posted Date: October 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2188729/v1

Abstract

Background

The topic of whether genetic screening for cancer risk should be implemented is complex. Using UK Biobank data, we 1) computed optimal risk thresholds for the detection of breast cancer, 2) examined the overlap of high-risk individuals identified by different risk predictors, and 3) evaluated the performance of risk predictor combinations.

Patients and methods

We studied 246,142 women without breast cancer at study entry. Risk predictors assessed include: the Gail model (GAIL), family history of breast cancer (FH, binary), 313-SNP breast cancer polygenic risk score (PRS), and carriership of loss-of-function variants in at least one of the 9 breast cancer predisposition genes (*ATM, BARD1, BRCA1, BRCA2, CHEK2, PALB2, RAD51D, RAD51C*, and *TP53*) (LoF). Absolute risk for developing invasive breast cancer was computed. Youden J-index was used to select optimal thresholds for defining high-risk.

Results

In total, 147,399 were considered at high risk for development of breast cancer within the next two years by at least one of the four breast cancer risk assessment tools examined ($Gail_{2-year>0.5\%}$: 47%, $PRS_{2-year>0.7\%}$: 30%, FH: 6%, and LoF: 1%); 92,851 (38%) were flagged by only one risk predictor. Seventy-nine percent of the breast cancers that did develop within the next two years were from the high-risk group. When compared to a random sample, the biggest gain in proportion of breast cancer cases was found within women at PRS high-risk, followed by GAIL, FH and LoF. The best-performing combinatorial model comprises a union of high-risk women identified by PRS, FH, and LoF (AUC_{2-year} [95% CI]: 62.2 [60.8 to 63.6]). Assigning individual weights to each risk prediction tool appeared to increase the discriminatory ability.

Conclusion

Our findings suggest that risk-based breast cancer screening may require a multi-pronged approach that includes PRS, breast cancer predisposition genes, family history, and other recognized risk factors.

Introduction

Breast cancer accounts for 15.5% of the 4,429,323 cancer deaths observed globally among women[1]. The most efficient way to identify breast cancer at an early stage and reduce mortality is through serial screening with mammography[2–6]. The most common target risk group in high-income countries offering population-based mammography screening is women between 50 to 69 years of age[7]. However, more than half of the 2,261,419 breast cancers diagnosed worldwide are outside this age group (< 50 years: 29.4%; ≥70 years: 22.4%) (GLOBOCAN, accessed Sep 19, 2022). The evidence regarding the degree of breast cancer mortality reduction in women under 50 or over 69, as well as the factors that determine the benefits and risks of mammography, such as the type of mammography (digital vs. screen-film), the number of views, and the interval between screenings, are inconsistent[3].

A more individualised, risk-based approach to breast cancer screening has garnered considerable lay public, health policy, and academic interest[8]. Evidence suggests that personalised risk-based screening, as opposed to one-size-fits-all age-based screening, improves the balance between the benefits and risks of breast cancer screening[9]. Knowing who to screen may reduce the negative outcomes of screening on a population level[10]. Individual needs may be used to guide and customise screening strategies, including starting age, stopping age, interval, and modality[11].

In an otherwise healthy population, breast cancer screening should ideally be done when the risk of the disease is high enough to offset the harms of overdiagnosis and overtreatment[5,12,13]. Women between the ages of 40 and 49 have a lower probability of developing breast cancer than older women, but the forms of breast cancer that do so are frequently more aggressive and have a worse prognosis[14]. Additionally, younger women are expected to live longer and have fewer comorbidities[14]. As for the older age group (≥ 70 years), there are doubts regarding the efficacy of mammography screening in reducing mortality due to a higher burden from non-cancer comorbidities[15–17]. Some have argued that older women may gain little from continuing routine screening mammography as the risk of overdiagnosis and unnecessary treatment may compromise quality of life and physical function[18]. “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 personalised risk assessments and lead to better patient outcomes[14].

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

- A higher risk of breast cancer exists in those who have a family history of the disease[19,20]. The elevated risk is likely brought on by genetic factors, but it might also be brought on by common lifestyle variables or other shared family traits[21].

- The Gail model, which incorporates classic breast cancer 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 breast cancer, is widely used and validated in many populations of different ancestry[22–27].
- Mammographic density, ascertained from the appearance of the breast tissue on mammograms, is another risk factor for breast cancer[28–30]. Studies have argued that mammographic density is more strongly associated with breast cancer risk than risk factors in the Gail model[31].
- It is possible to predict whether a person's risk of developing breast cancer is caused by a particular genetic vulnerability (such rare loss-of-function variants in *BRCA1* or *BRCA2*)[32–35]. Absolute lifetime risks of developing breast cancer associated with *BRCA1/2* mutations have been reported to be ~ 80%, much higher than the ~ 12% typically observed in the general population[36]. Large consortia efforts have also identified other clinically useful breast cancer predisposition genes (*ATM, CHEK2, PALB2, BARD1, RAD51C, RAD51D*, and *TP53*)[37].
- Polygenic risk scores (PRS) that sums up the effects of multiple common variants associated with breast cancer have been implemented in pilot precision health initiatives that stratifies individuals by their personal disease risk[38–47].

With the continuous development and refinement of breast cancer risk assessment tools, “How much does genetics add?” becomes of interest. There is increasing evidence that combining genetic data with standard risk instruments meaningfully enhances risk stratification and improves discriminatory value in mammography screening programmes[48,49]. However, the extent of overlap between the high-risk individuals identified by different tools is unclear. In this study, we use the large UK Biobank dataset comprising 246,142 women to assess the proportion of women identified at high-risk of developing breast cancer ascertained by different means and evaluate the proportion of high-risk individuals who eventually developed the malignancy.

Methods

Study population

The UK Biobank is a publicly available scientific database and research tool which has curated comprehensive genetic and health data from ~ 500,000 individuals in the United Kingdom [50](10.1371/journal.pmed.1001779). 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 one of 22 evaluation centres in England, Scotland, and Wales were requirements for participants. A touch screen survey, a nurse interview, and physical measures were all completed by participants. Additionally, participants agreed for data to be linked to primary care, hospital episode data, and death registers. Our cohort was restricted to 264,741 female participants (application 86846) (**Supplementary Fig. 1**).

Breast cancer polygenic risk score (PRS)

Genome-wide genetic data is 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[51]:

“Genotype calling was performed by Affymetrix (now part of ThermoFisher Scientific) on two closely related purpose-designed arrays. ~50,000 participants were run on the UK BiLEVE Axiom array (Resource 149600) and the remaining ~ 450,000 were run on the UK Biobank Axiom array (Resource 149601). The dataset combines results from both arrays (see Data-Field 22000) and there are 805,426 markers in the released genotype data. The positions of markers in the data are in GRCh37 coordinates. It was not possible to assay genotypes for some participants (~ 3%) as sufficient DNA could not be extracted from their blood samples.

The genotype data were quality controlled (QC). In addition, the dataset was phased and ~ 96M genotypes were imputed using computationally efficient methods combined with the Haplotype Reference Consortium and UK10K haplotype resources. Variants are stored in the compressed and indexed BGENv1.2 format. The imputed genotypes are aligned to the + strand of the reference and the positions are in GRCh37 coordinates.

Information from the QC pipeline, such as array, and important genetic properties of the data such as population structure and relatedness are available. Details of these analyses, and the methods used to derive other data such as imputation and haplotypes, are given in Bycroft et al[52].”

A list of the 313 SNPs and associated weights included in the breast cancer PRS is given in **Supplementary Table 1** [38]. The plink code used to extract 313 SNPs for building the breast cancer PRS (using chr21 as example) is: “plink2 –bgen ukb22828_c21_b0_v3.bgen –sample ukb22828_c21_b0_v3_s487201.sample –keep-females –remove-nosex –extract temp$1.bim –make-bed –out ukb22828_c21_b0_v3”. In the extraction, 308 variants and 264,246 females pass filters and QC (total genotyping rate = 0.977395).

Carriers of loss-of-function (LoF) variants in nine breast cancer risk genes

The WES 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)[53]. The Swiss Army Knife (v4.8.0) executable was used to run PLINK2[54].

Poor quality variants were excluded with the provided helper file (ukb23158_500k_OQFE.90pct10dp_qc_variants.txt), which is a single-column text file containing variants failing the “90pct10dp” depth filter in the CHR:POS:REF:ALT format. 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 (i.e. DP > = 10).

The file ukb23158_500k_OQFE.annotations.txt.gz provided by the UK Biobank defines a functional annotation for each variant given a gene set. Variants annotated with missense(0/5), missense(5/5), missense(> = 1/5), or synonymous were excluded. LoF variants (n = 795) in nine breast cancer risk genes (*ATM* [ENSG00000149311], n = 222; *BRCA1* [ENSG00000012048], n = 115; *BRCA2* [ENSG00000139618], n = 179; *CHEK2* [ENSG00000183765], n = 60; *PALB2* [ENSG00000083093], n = 85; *BARD1* [ENSG00000138376], n = 49; *RAD51C* [ENSG00000108384], n = 23; *RAD51D* [ENSG00000185379], n = 47; or *TP53* [ENSG00000141510], n = 15) were extracted (10.1056/NEJMoa1913948). The max minor allele frequency was set at 0.01.

The command line option (using chr21 as an example) to extract LoF variants was "plink2 –bfile ukb23158_c21_b0_v1 –no-psam-pheno –exclude ukb23158_500k_OQFE.90pct10dp_qc_variants.txt –extract ukb23158_500k_OQFE.annotations-9genes-795LoF –keep-females –max-maf 0.01 –make-bed –out chr21".

The resulting WES dataset included 508 LoF variants with at least one carrier 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), family history of breast cancer (Data-Fields: 20110 for mother and 20111 for siblings), menopausal status (Data-Field: 2724). A maximum value of 2 is possible for family history of breast cancer, where the individual's mother and at least one sibling had breast cancer. Information on ever had breast cancer screening visits (Data-Field: 2674) was also retrieved.

Breast cancer case ascertainment

Invasive breast cancer was determined using the 9th (174*) and 10th (C50*) versions of the International Classification of Diseases (Data-Fields 40013 and 40006, 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 breast cancers were diagnosed at age 80 and above. Due to the late age at diagnosis, these cases were considered as non-case in our analysis. In situ breast cancer cases were included as non-cases.

A total of 253,953 females had information on both PRS and LoF (**Supplementary Fig. 1**). Prevalent breast cancer cases (n = 7,811) were excluded. The resulting analytical cohort consisted of 246,142 females, with 7,620 incident cases of breast cancer (latest case diagnosed in year 2020).

Statistical analysis

Associations between risk factors of interest and invasive breast cancer diagnosis before age 80 years were tested using Chi-square test for categorical variables and Kruskal-Wallis test for continuous variables.

Our outcomes of interest are invasive breast cancer diagnosis within 2-, 5-, and 10-years post-study entry and lifetime (before age 80 years). For each period of interest, the corresponding x-year absolute risk (2-, 5-, 10-year, and lifetime cumulative risk [before age 80 years]) was computed.

The x-year absolute risk (2-year, 5-year, 10-year, and lifetime at age 80 years) was predicted using the package (*BCRA)* for the Gail model[22,55]. The method to obtain the x-year absolute risk computed from PRS is described previously[39]. Breast cancer incidence rates from 2011 to 2015, and mortality rates of 2016, were used in the PRS absolute risk calculation[56,57]. The distributions of x-year absolute risks predicted by PRS and the Gail model are illustrated in **Supplementary Fig. 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-ndex (from pROC package in R)[58,59]. This threshold optimization was repeated for the x-year absolute risks computed from the Gail model. Family history of breast cancer 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, 2 + were adopted. Carriers of LoF variants in at least one of the 9 breast cancer predisposition genes (*ATM, BARD1, BRCA1, BRCA2, CHEK2, PALB2, RAD51D, RAD51C*, and *TP53*) were considered LoF variant carriers.

The women identified to be at high risk of developing breast cancer according to the four risk prediction tools (PRS, Gail, family history, 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 four risk prediction tools on breast cancer risk (**Additional File 1**).

R version 4.0.3 was used in all analyses.

Results

Study population characteristics

We studied 246,142 women without breast cancer at study entry. The median age was 56 years (interquartile range [IQR]: 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 breast cancer screening (n =

194,441, 79%). Six percent (n = 15,304) and < 1% (n = 466) reported one (mother or sibling) and two (mother and at least one sibling) first-degree family history of breast cancer, respectively. As of year 2020, 3% of the study population (n = 7,620) developed invasive breast cancer. The median age at diagnosis was 63 years (IQR: 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 due to the large sample size (Table 1).

Table 1
Characteristics of study participants. Column percentages are shown within brackets. IQR: interquartile range; P: p-value from Chi-square test (categorical variable) or Kruskal-Wallis test (continuous variable).

		Breast cancer diagnosis before age 80 years		
	All	No	Incident	P
	N = 246,142	n = 238,522	n = 7,620	
Median age at recruitment (IQR)	56 (50 to 63)	56 (50 to 63)	57 (51 to 63)	< 0.001
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%)	7,275 (95%)	< 0.001
African American (includes African, Caribbean, Black or Black British, and other Black background)	4,092	4,017 (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)	4,092	3,971 (2%)	121 (2%)	
Other (includes mixed, White and Black, White and Black African, White and Asian, unknown, and prefer not to answer)	4,953	4,823 (2%)	130 (2%)	
First degree family history of breast cancer (mother and siblings)				
None	230,372	223,519 (94%)	6,853 (90%)	< 0.001
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%)	2,596 (34%)	0.009
12 to 13	103,892	100,562 (42%)	3,330 (44%)	
≤ 11	47,014	45,534 (19%)	1,480 (19%)	
Unknown	7,659	7,445 (3%)	214 (3%)	
Number of children				
0	45,798	44,291 (19%)	1,507 (20%)	0.002
1	32,833	31,765 (13%)	1,068 (14%)	
2	107,513	104,218 (44%)	3,295 (43%)	
3+	59,442	57,707 (24%)	1,735 (23%)	
Unknown	556	541 (0%)	15 (0%)	
Age at first live birth, years				
No child	45,798	44,291 (19%)	1,507 (20%)	< 0.001
≤ 19	21,460	20,905 (9%)	555 (7%)	
20 to 24	63,671	61,831 (26%)	1,840 (24%)	
25 to 29	70,749	68,525 (29%)	2,224 (29%)	
≥ 30	43,513	42,044 (18%)	1,469 (19%)	
Unknown	951	926 (0%)	25 (0%)	
Menopausal status				
Yes	147,683	142,940 (60%)	4,743 (62%)	< 0.001
No	59,266	57,586 (24%)	1,680 (22%)	

		Breast cancer diagnosis before age 80 years		
Unknown	39,193	37,996 (16%)	1,197 (16%)	
Ever attended breast cancer screening				
Yes	194,441	188,111 (79%)	6,330 (83%)	< 0.001
No	50,979	49,707 (21%)	1,272 (17%)	
Unknow	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)	< 0.001
High-penetrance breast cancer genes (ATM, BARD1, BRCA1, BRCA2, CHEK2, PALB2, RAD51D, RAD51C, TP53**)**				
No	243,137	235,728 (99%)	7,409 (97%)	< 0.001
Yes (at least one loss-of-function variant)	3,005	2,794 (1%)	211 (3%)	

Optimal thresholds for the definition of high-risk

The Youden J-Index is frequently used to summarize the receiver operating characteristic curve [59]. In this study, we used it to assess the efficacy of a diagnostic marker and to select the best threshold value or cut-off point for PRS and Gail model-computed absolute risks. The most optimal Youden J-Index was achieved with absolute risk cut-offs of 0.7% (PRS_{2-year} AUC [95%CI]: 65.1 [63.6 to 66.7]) and 0.5% ($Gail_{2-year}$ AUC [95%CI]: 59.3 [57.7 to 60.9]). **Supplementary Fig. 3** summarizes the corresponding AUC values when considering 2-, 5-, 10-years and lifetime absolute risks. The most optimal cut-offs were used for subsequent analyses.

Overlap of high-risk individuals identified by different breast cancer risk assessment tools

The proportion of women flagged as high risk by $Gail_{2-year>0.5\%}$, $PRS_{2-year>0.7\%}$, FH, and LoF were 47% (n = 115,986), 30% (n = 73,775), 6% (n = 15,770), and 1% (n = 3,005), respectively (Fig. 1). **Supplementary Fig. 4** shows corresponding Venn diagrams for 5-year, 10-year, and lifetime absolute risks. Thirty-eight percent of the 246,142 women in the study were considered high-risk by only one risk prediction tool ($PRS_{2-year>0.5\%}$ unique individuals = 28,630, $Gail_{2-year>0.7\%}$ unique individuals = 61,911, FH unique individuals = 1,142, LoF unique individuals = 1,168).

Seventy-nine percent of the 1,209 breast cancer cases that developed within two years were identified to be at high risk by at least one of the four breast cancer predictors examined. Using a two-year absolute risk cut-off 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 breast cancer cases that developed within the next two years of assessment; LoF carriers made up less than 4% (n = 44).

Improvement in the number of breast cancers identified within a high-risk group vs a random sample

Figure 2 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, 10 years or lifetime, y-axis). When considered as single risk predictors, breast cancer PRS was associated with the highest gain in 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 carriership of LoF variants in high-penetrant breast cancer genes. The best-performing combinatorial model comprises PRS, FH, and LoF (AUC_{2-year} [95% CI]: 62.2 [60.8 to 63.6]) (Fig. 2, Table 2).

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 Fig. 2. *x-year absolute risk threshold to define high risk. PRS: breast cancer polygenic risk score; GAIL: the Gail model; LoF: carrier of loss-of-function variants in high-penetrance breast cancer genes; AUC: area under the receiver operating curve; FPR: false positive rate; FNR: false negative rate; TPR: true positive rate; TNR: true negative rate.

						95% CI of AUC					
x-year absolute risk	**PRS***	**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
Lifetime at age 80 years	4.2	PRS ⋃ FH ⋃ LoF	55.3	61.8	58.5	58.0	59.1	95.6	2.3	4.4	97.7

Assigning weights to each risk prediction tool improves discriminatory ability

We next examined whether the discriminatory ability changes when risk prediction tools were assigned different weights (Table 3). This model would account for effect overlap between the different tools. The best-performing backward stepwise logistic regression model retained all four risk prediction tools (Youden J-Index = 24.7, AUC_{2-year} [95% CI]: 66.4 [64.8 to 67.9]). After manual removal of the Gail model, the AUC_{2-year} achieved was 66.3 (95% CI: 64.7 to 67.8).

Table 3

Discriminatory ability and performance measures of different strategies in identifying high-risk women. The models show beta weights from a stepwise logistic regression with backward removal predicting breast cancer risk using the four different risk prediction tools. p: probability from the combinatorial model; PRS: polygenic risk scores; GAIL: the Gail model; LoF: carrier of loss-of-function variants in high-penetrance breast cancer genes; AUC: area under the receiver operating curve; FPR: false positive rate; FNR: false negative rate; TPR: true positive rate; TNR: true negative rate.

			Model statistics											
	Absolute risk cut-off (%)								95% CI of AUC					
x-year absolute risk	PRS	GAIL	Model	Threshold (p)	Sensitivity	Specificity	Youden J-statistics	AUC	lower	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	0.5	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	0.9	2	99.1
10-year	2.9	3	Logit(p): -4.43 + 0.15 PRS + 0.10 GAIL + 0.16 FH + 0.85 LoF	2.5	59	60.6	19.6	63.5	62.9	64.2	96	1.8	4	98.2
Lifetime at age 80 years	4.2	9.5	Logit(p): -4.42 + 0.10 PRS + 0.07 GAIL + 0.04 FH + 0.81 LoF	3	53.4	65.3	18.7	62.5	61.9	63.2	95.3	2.2	4.7	97.8
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

			Model statistics										
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
Lifetime at age 80 years	4.2	-	Logit(p): -3.86 + 0.08 PRS + 0.44 FH + 0.82 LoF	2.9	54.9	62.1	17.1	61.4	60.8	62.1	95.6	2.3	4.4	97.7

Discussion

Stratifying population-level service users by their individual breast cancer risk may improve resource utilization and alleviate the issue of overdiagnosis[60]. We identified individuals at high risk of developing breast cancer 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 breast cancer absolute risks were computed. The analysis associated with two-year breast cancer 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 development of breast cancer within the next two years by at least one of the four breast cancer risk assessment tools examined. Among the high-risk individuals, 92,851 (38%) were flagged by only one risk predictor. Seventy-nine percent of the breast cancers that did develop within the next two 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 breast cancer cases detected when compared to 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 one risk assessment tool suggests that breast cancer screening may benefit from a multi-prong approach including genetic and non-genetic risk factors. Numerous studies have shown that breast cancer PRS exerts an effect distinct from traditional risk factors[61]. The effect of PRS on breast cancer risk is known not to be correlated with family history[38,41,62–64]. PRSs have also been shown to be largely independent of other known risk factors for breast cancer, such as mammographic density[65], lifestyle factors[62,66–68], reproductive factors and hormone use[62,68]. In a European study comprising using 72,284 cases and 80,354 controls from the Breast Cancer Association Consortium, Kapoor et al. showed that the effects of PRS did not differ across different strata of conventional risk factors[66].

Considering different risk prediction models in tandem improves performance. In a study comprising 126,894 women, Yang et al. showed women at the highest risk of developing breast cancer (top 5th percentile) estimated by PRS and non-genetic risk factors 3.84- and 2.10-times more likely to develop the disease compared to the average risk group (40th to 60th percentile)[48]. While the non-genetic risk score had limited predictive ability by itself, the joint predictive model performed better than PRS or non-genetic risk score alone[48]. In the FinnGen study (n = 122,978, 8,401 breast cancer cases), authors showed that PRS improves risk prediction in women with first-degree family members who have been diagnosed with breast cancer[69]. In the Predicting the Risk of Cancer At Screening (PROCAS study), Evans et al. evaluated the risk prediction capacities of traditional risk factors, mammographic density, PRS, and a gene panel[49]. The authors found that whilst PRS improved risk stratification significantly when compared to a model based only on mammographic density and conventional risk factors, the inclusion of gene panels showed no appreciable effect. In our study, the best discriminatory ability was associated with a combinatorial model including PRS, FH, and LoF.

We show that risk stratification using PRS, FH and LoF offered best gain in terms of the number of breast cancer cases detected in a high-risk population when compared to 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) as compared with taking a random sample of the same proportion. This result is in agreement with previous works examining the impact of using multiple risk prediction tools in risk-based breast cancer screening scenarios. Darabi et al estimated that a customised screening strategy with input from multiple risk models (e.g. conventional risk factors, mammographic density, and polygenic risk scores) captures 10% more cases than an age-based approach[70]. For every thousand 50-year-old women in Nurses' Health Study, the Gail model alone identified 2 individuals associated with double the average five-year absolute risk (2.27%)[71]. However, the addition of PRS and mammographic density and exogenous hormone use identified 6.6% at elevated risk[71]. In another study, Hsieh et al showed that PRS enhanced the discriminatory ability of the non-genetic model and flagged more breast cancer cases as high-risk[72]. In a scenario where breast cancer 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[73].

An issue that emerged from this analysis is the large number of women considered high-risk by PRS and the Gail model. All breast cancer cases would be detected if all women were considered high risk, which defeats the aim of risk stratification. We thus assigned weights to individual risk assessment tools and found that the discriminatory value improved (95%CI of the AUCs did not overlap). However, the threshold for what is considered high risk will

ultimately depend on the healthcare resources available in each country. In addition, randomised trials, preferably with breast cancer mortality as the outcome, would be needed to establish the benefit of risk stratification.

Another issue is that many women who are considered at high risk may never experience the disease, and many tumors show no alterations in known breast cancer genes[38,74]. Each breast-cancer-associated common variant typically accounts for far less than 1% of an individual's risk. The culmination of GWAS efforts in identifying common variants to date explains no more than 40% of the two-fold familial relative risk for invasive breast cancer[75]. Twenty-one percent of breast cancer cases in this study were not flagged as high risk by any of the four risk assessment tools examined.

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 discounted[76]. Participants in the UK Biobank are known to be of higher economic status and to have fewer risk factors related to lifestyle[77]. 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[76]. 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 family history, number of breast biopsies, and history of atypical hyperplasia may explain the poor performance of the Gail model in breast cancer risk stratification. Mammographic density, a strong risk factor for breast cancer, is also not available as a variable for breast cancer risk assessment. As information on breast cancer stage and hormone-receptor subtype is not available in the UK Biobank, we were unable to subset the analyses by tumor features.

Our findings suggest that risk-based breast cancer screening programmes may benefit from a multi-pronged approach that includes PRS, pathogenic mutations in breast cancer predisposition genes, family history, and other recognized risk factors. Nonetheless, to be successful, screening programs require significant health resources, a strong infrastructure, and capability within the country's health care system[78]. 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 breast cancer is realised.

Declarations

ADDITIONAL FILES

Additional file 1 - R codes.pdf

Additional file 2 - Supporting data for Figure 2 getting best combination.xlsx

DATA AVAILABILITY STATEMENT

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

ETHICAL CONSIDERATIONS

UK Biobank has approval from the NHS 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 analyse 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).

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 CONTRIBUTIONS

All authors have made substantial contributions to conception and design, involved in drafting the manuscript or revising it critically for important intellectual content, and given final approval of the version to be published. All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

COMPETING INTERESTS

None declared.

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin **71**, 209–249, doi:10.3322/caac.21660 (2021).
2. de Koning, H. J. Mammographic screening: evidence from randomised controlled trials. Ann Oncol **14**, 1185–1189, doi:10.1093/annonc/mdg319 (2003).
3. Mandrik, O. *et al.* Systematic reviews as a 'lens of evidence': Determinants of benefits and harms of breast cancer screening. Int J Cancer **145**, 994–1006, doi:10.1002/ijc.32211 (2019).
4. Coleman, C. Early Detection and Screening for Breast Cancer. Semin Oncol Nurs **33**, 141–155, doi:10.1016/j.soncn.2017.02.009 (2017).
5. Bleyer, A. & Welch, H. G. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med **367**, 1998–2005, doi:10.1056/NEJMoa1206809 (2012).
6. World Health, O. *WHO position paper on mammography screening*. 78 p. (World Health Organization, 2014).
7. Ebell, M. H., Thai, T. N. & Royalty, K. J. Cancer screening recommendations: an international comparison of high income countries. Public Health Rev **39**, 7, doi:10.1186/s40985-018-0080-0 (2018).
8. Sud, A., Turnbull, C. & Houlston, R. Will polygenic risk scores for cancer ever be clinically useful? NPJ Precis Oncol **5**, 40, doi:10.1038/s41698-021-00176-1 (2021).
9. Kerlikowske, K. & Bibbins-Domingo, K. Toward Risk-Based Breast Cancer Screening. Ann Intern Med **174**, 710–711, doi:10.7326/M21-0398 (2021).
10. Gilbert, N. The pros and cons of screening. Nature **579**, S2-S4, doi:10.1038/d41586-020-00841-8 (2020).
11. Shieh, Y. *et al.* Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. J Natl Cancer Inst **109**, doi:10.1093/jnci/djw290 (2017).
12. Esserman, L. J. *et al.* Addressing overdiagnosis and overtreatment in cancer: a prescription for change. Lancet Oncol **15**, e234-242, doi:10.1016/S1470-2045(13)70598-9 (2014).
13. Kowalski, A. E. Mammograms and Mortality: How Has the Evidence Evolved? J Econ Perspect **35**, 119–140, doi:10.1257/jep.35.2.119 (2021).
14. Grimm, L. J., Avery, C. S., Hendrick, E. & Baker, J. A. Benefits and Risks of Mammography Screening in Women Ages 40 to 49 Years. J Prim Care Community Health **13**, 21501327211058322, doi:10.1177/21501327211058322 (2022).
15. Richman, I. B. & Gross, C. P. New Breast Cancer Screening Technologies in Older Women-Is It Time to Pump the Brakes? JAMA Intern Med **179**, 289–290, doi:10.1001/jamainternmed.2018.7767 (2019).
16. Demb, J. *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 Epidemiology, Biomarkers & Prevention **27**, 358–358, doi:10.1158/1055-9965.EPI-18-0065 (2018).
17. Brawley, O. W. On Breast Cancer Screening in Older Women. Ann Intern Med **175**, 127–128, doi:10.7326/M21-4235 (2022).
18. de Glas, N. A. *et al.* Effect of implementation of the mass breast cancer screening programme in older women in the Netherlands: population based study. BMJ **349**, g5410, doi:10.1136/bmj.g5410 (2014).
19. Cole, P. What should the physician ask? *Cancer* **37**, 434–436, doi:https://doi.org/10.1002/1097-0142(197601)37:1+<434::AID-CNCR2820370706>3.0.CO;2-O (1976).
20. Haagensen, C. D. Family History of Breast Carcinoma in Women Predisposed to Develop Breast Carcinoma1. JNCI: Journal of the National Cancer Institute **48**, 1025–1027, doi:10.1093/jnci/48.4.1025 (1972).
21. Moller, S. *et al.* The Heritability of Breast Cancer among Women in the Nordic Twin Study of Cancer. Cancer Epidemiol Biomarkers Prev **25**, 145–150, doi:10.1158/1055-9965.EPI-15-0913 (2016).
22. Gail, M. H. *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst **81**, 1879–1886, doi:10.1093/jnci/81.24.1879 (1989).

23. Bondy, M. L. & Newman, L. A. Assessing breast cancer risk: evolution of the Gail Model. J Natl Cancer Inst **98**, 1172–1173, doi:10.1093/jnci/djj365 (2006).
24. Costantino, J. P. *et al.* Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst **91**, 1541–1548, doi:10.1093/jnci/91.18.1541 (1999).
25. Rockhill, B., Spiegelman, D., Byrne, C., Hunter, D. J. & Colditz, G. A. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. J Natl Cancer Inst **93**, 358–366, doi:10.1093/jnci/93.5.358 (2001).
26. Ho, P. J. *et al.* Breast cancer risk stratification for mammographic screening: A nation-wide screening cohort of 24,431 women in Singapore. Cancer Med **10**, 8182–8191, doi:10.1002/cam4.4297 (2021).
27. Bondy, M. L. & Newman, L. A. Breast cancer risk assessment models: applicability to African-American women. Cancer **97**, 230–235, doi:10.1002/cncr.11018 (2003).
28. Boyd, N. F. *et al.* Mammographic density: a heritable risk factor for breast cancer. Methods Mol Biol **472**, 343–360, doi:10.1007/978-1-60327-492-0_15 (2009).
29. Boyd, N. F. *et al.* Mammographic breast density as an intermediate phenotype for breast cancer. The Lancet Oncology **6**, 798–808, doi:https://doi.org/10.1016/S1470-2045(05)70390-9 (2005).
30. Pettersson, A. *et al.* Mammographic density phenotypes and risk of breast cancer: a meta-analysis. J Natl Cancer Inst **106**, doi:10.1093/jnci/dju078 (2014).
31. Boyd, N. F., Martin, L. J., Yaffe, M. J. & Minkin, S. Mammographic density and breast cancer risk: current understanding and future prospects. Breast Cancer Res **13**, 223, doi:10.1186/bcr2942 (2011).
32. Ang, B. H. *et al.* Predicting the Likelihood of Carrying a BRCA1 or BRCA2 Mutation in Asian Patients With Breast Cancer. J Clin Oncol **40**, 1542–1551, doi:10.1200/JCO.21.01647 (2022).
33. Domchek, S. M. *et al.* Application of breast cancer risk prediction models in clinical practice. J Clin Oncol **21**, 593–601, doi:10.1200/JCO.2003.07.007 (2003).
34. Barcenas, C. H. *et al.* Assessing BRCA carrier probabilities in extended families. J Clin Oncol **24**, 354–360, doi:10.1200/JCO.2005.02.2368 (2006).
35. Lee, A. *et al.* BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med **21**, 1708–1718, doi:10.1038/s41436-018-0406-9 (2019).
36. Kuchenbaecker, K. B. *et al.* Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA **317**, 2402–2416, doi:10.1001/jama.2017.7112 (2017).
37. Breast Cancer Association, C. *et al.* Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med **384**, 428–439, doi:10.1056/NEJMoa1913948 (2021).
38. Mavaddat, N. *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet **104**, 21–34, doi:10.1016/j.ajhg.2018.11.002 (2019).
39. Ho, W. K. *et al.* European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. Nat Commun **11**, 3833, doi:10.1038/s41467-020-17680-w (2020).
40. Ho, W. K. *et al.* Polygenic risk scores for prediction of breast cancer risk in Asian populations. Genet Med **24**, 586–600, doi:10.1016/j.gim.2021.11.008 (2022).
41. Mavaddat, N. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst **107**, doi:10.1093/jnci/djv036 (2015).
42. Liu, J. *et al.* BREAst screening Tailored for HEr (BREATHE)-A study protocol on personalised risk-based breast cancer screening programme. PLoS One **17**, e0265965, doi:10.1371/journal.pone.0265965 (2022).
43. Roux, A. *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) randomised clinical trial. BMC Cancer **22**, 507, doi:10.1186/s12885-022-09484-6 (2022).
44. Brooks, J. D. *et al.* Personalized Risk Assessment for Prevention and Early Detection of Breast Cancer: Integration and Implementation (PERSPECTIVE I&I). J Pers Med **11**, doi:10.3390/jpm11060511 (2021).
45. Gabrielson, M. *et al.* Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). Int J Epidemiol **46**, 1740–1741g, doi:10.1093/ije/dyw357 (2017).
46. Esserman, L. J., Study, W. & Athena, I. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer **3**, 34, doi:10.1038/s41523-017-0035-5 (2017).
47. Keane, H. *et al.* Development and pilot of an online, personalized risk assessment tool for a breast cancer precision medicine trial. NPJ Breast Cancer **7**, 78, doi:10.1038/s41523-021-00288-8 (2021).
48. Yang, Y. *et al.* Incorporating Polygenic Risk Scores and Nongenetic Risk Factors for Breast Cancer Risk Prediction Among Asian Women. JAMA Netw Open **5**, e2149030, doi:10.1001/jamanetworkopen.2021.49030 (2022).

49. Evans, D. G. R. *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 **24**, 1485–1494, doi:10.1016/j.gim.2022.03.009 (2022).
50. Sudlow, C. *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 **12**, e1001779, doi:10.1371/journal.pmed.1001779 (2015).
51. Bycroft, C. *et al.* Genome-wide genetic data on ~ 500,000 UK Biobank participants. *bioRxiv*, 166298, doi:10.1101/166298 (2017).
52. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. Nature **562**, 203–209, doi:10.1038/s41586-018-0579-z (2018).
53. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. Nature **599**, 628–634, doi:10.1038/s41586-021-04103-z (2021).
54. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).
55. Banegas, M. P. *et al.* Projecting Individualized Absolute Invasive Breast Cancer Risk in US Hispanic Women. J Natl Cancer Inst **109**, doi:10.1093/jnci/djw215 (2017).
56. *Cancer incidence rates for combined years by: site(third digit), sex and age, England, 2011–2015*, <www.ons.gov.uk> (
57. *Incidence of, and mortality from breast cancer in England, 2004 to 2017.*, <www.ons.gov.uk> (
58. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35, doi:https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3 (1950).
59. Schisterman, E. F., Faraggi, D., Reiser, B. & Hu, J. Youden Index and the optimal threshold for markers with mass at zero. Stat Med **27**, 297–315, doi:10.1002/sim.2993 (2008).
60. Gierach, G. L., Choudhury, P. P. & Garcia-Closas, M. Toward Risk-Stratified Breast Cancer Screening: Considerations for Changes in Screening Guidelines. JAMA Oncol **6**, 31–33, doi:10.1001/jamaoncol.2019.3820 (2020).
61. Ho, P. J. *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 **20**, 150, doi:10.1186/s12916-022-02334-z (2022).
62. Hou, C. *et al.* Development and validation of polygenic risk scores for prediction of breast cancer and breast cancer subtypes in Chinese women. BMC Cancer **22**, 374, doi:10.1186/s12885-022-09425-3 (2022).
63. Li, H. *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 **19**, 30–35, doi:10.1038/gim.2016.43 (2017).
64. Wang, Y., Zhu, M., Ma, H. & Shen, H. Polygenic risk scores: the future of cancer risk prediction, screening, and precision prevention. **1**, 129–149, doi:doi:10.1515/mr-2021-0025 (2021).
65. van Veen, E. M. *et al.* Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. JAMA Oncol **4**, 476–482, doi:10.1001/jamaoncol.2017.4881 (2018).
66. Kapoor, P. M. *et al.* Combined Associations of a Polygenic Risk Score and Classical Risk Factors With Breast Cancer Risk. J Natl Cancer Inst **113**, 329–337, doi:10.1093/jnci/djaa056 (2021).
67. Arthur, R. S., Wang, T., Xue, X., Kamensky, V. & Rohan, T. E. Genetic Factors, Adherence to Healthy Lifestyle Behavior, and Risk of Invasive Breast Cancer Among Women in the UK Biobank. J Natl Cancer Inst **112**, 893–901, doi:10.1093/jnci/djz241 (2020).
68. Rudolph, A. *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 **47**, 526–536, doi:10.1093/ije/dyx242 (2018).
69. Mars, N. *et al.* The role of polygenic risk and susceptibility genes in breast cancer over the course of life. Nat Commun **11**, 6383, doi:10.1038/s41467-020-19966-5 (2020).
70. Darabi, H. *et al.* Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. Breast Cancer Res **14**, R25, doi:10.1186/bcr3110 (2012).
71. Zhang, X. *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 **15**, e1002644, doi:10.1371/journal.pmed.1002644 (2018).
72. Shieh, Y. *et al.* Breast cancer risk prediction using a clinical risk model and polygenic risk score. Breast Cancer Res Treat **159**, 513–525, doi:10.1007/s10549-016-3953-2 (2016).
73. Pashayan, N. *et al.* Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. Br J Cancer **104**, 1656–1663, doi:10.1038/bjc.2011.118 (2011).
74. Versteeg, R. Cancer: Tumours outside the mutation box. Nature **506**, 438–439, doi:10.1038/nature13061 (2014).
75. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet **52**, 572–581, doi:10.1038/s41588-020-0609-2 (2020).
76. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol **186**, 1026–1034, doi:10.1093/aje/kwx246 (2017).
77. Hanlon, P. *et al.* Associations between multimorbidity and adverse health outcomes in UK Biobank and the SAIL Databank: A comparison of longitudinal cohort studies. PLoS Med **19**, e1003931, doi:10.1371/journal.pmed.1003931 (2022).

78. Wedderburn, S. & McVeigh, T. P. Should All Individuals Be Screened for Genetic Predisposition to Cancer? *Genet Res (Camb)* 2021, 6611963, doi:10.1155/2021/6611963 (2021).

Figures

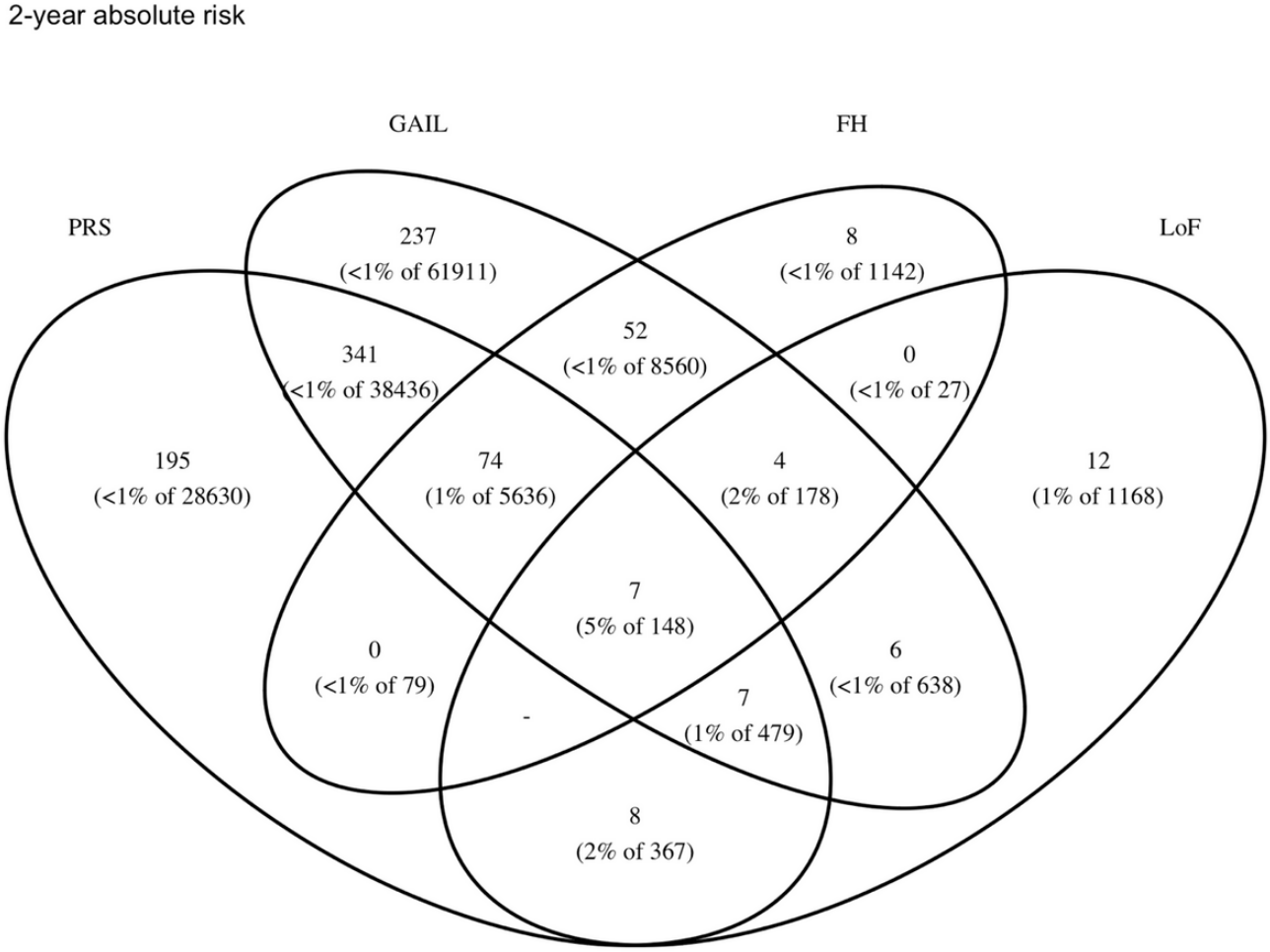

Figure 1

Number of incident invasive breast cancer events by the four 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: (i) x-year absolute risk above threshold as predicted by polygenic risk score (PRS: >0.7% for 2-year absolute risk, respectively), (ii) x-year absolute risk above threshold as predicted by the Gail model (GAIL: >0.5% for 2-year absolute risk, respectively), (iii) family history of breast cancer (yes, FH), and (iv) carriers of 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*).

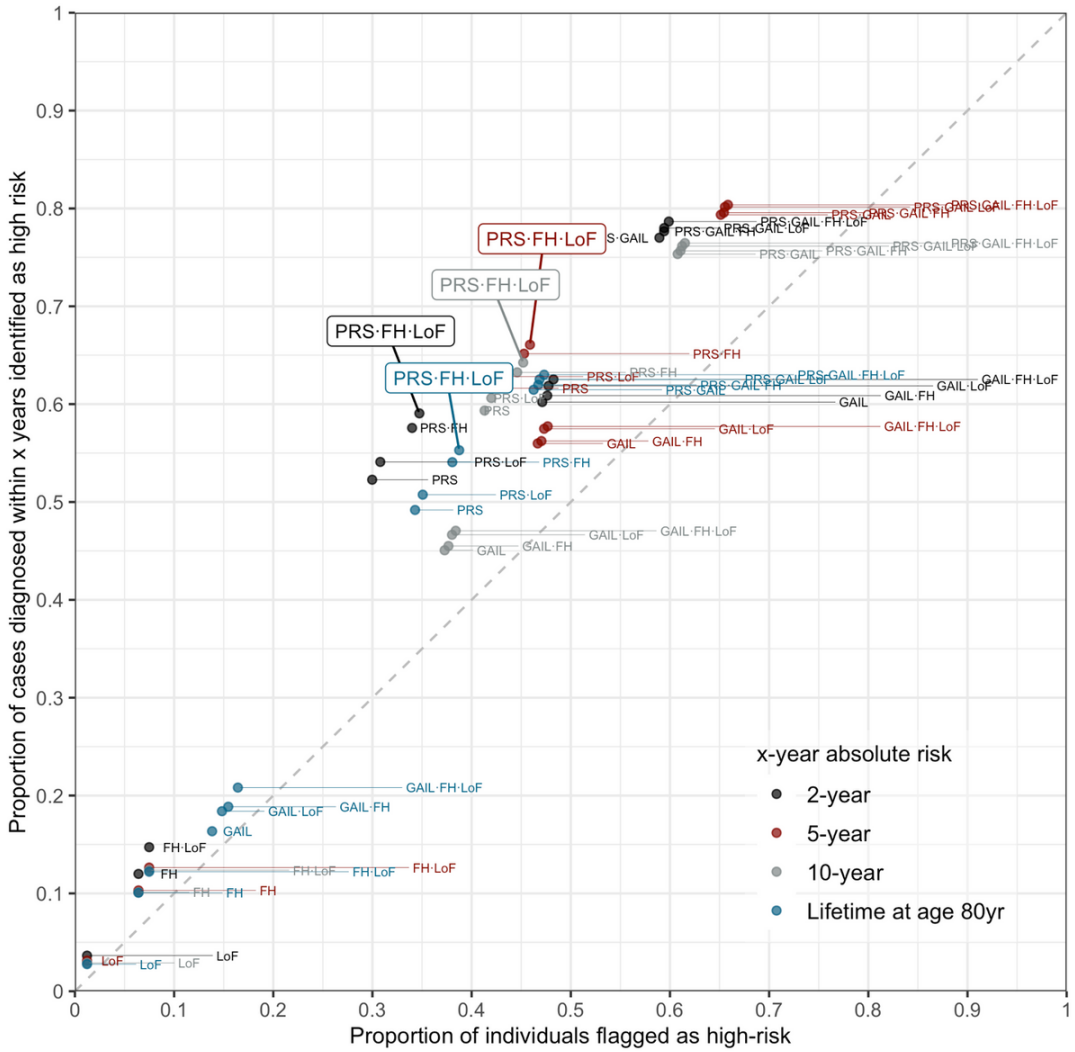

Figure 2

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, 10 or lifetime, y-axis). Breast cancer polygenic risk score (PRS) was associated with the highest gain in 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 carriership of loss-of-function variants in high-penetrant breast cancer genes (LoF). The best-performing combinatorial model (in boxed labels) comprises PRS, FH, and LoF.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1Rcodes.pdf
- Additionalfile2SupportingdataforFigure2gettingbestcombination.xlsx
- SupplementaryTablesandFigures.xlsx