Cancer Epidemiology 94 (2025) 102701

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

Utilizing polygenic risk score for breast cancer risk prediction in a Taiwanese population

Yi-Hsuan Lin [a], Chih-Chiang Hung [a,b,c], Guan-Cheng Lin [d], I.-Chen Tsai [a,e], Chih Yean Lum [a,*], Tzu-Hung Hsiao [c,f,g,**]

[a] Division of Breast Surgery, Department of Surgery, Taichung Veterans General Hospital, Taichung 40705, Taiwan
[b] Department of Applied Cosmetology, College of Human Science and Social Innovation, Hung Kuang University, Taichung 43302, Taiwan
[c] Ph.D Program in Translational Medicine, College of Life Sciences, National Chung Hsing University, Taichung 40227, Taiwan
[d] Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan
[e] College of Biomedical, China Medical University, Taichung, Taiwan
[f] Department of Public Health, Fu Jen Catholic University, New Taipei City 24205, Taiwan
[g] Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung 4022, Taiwan

ARTICLE INFO

Keywords:
Breast cancer risk
PGS000508
Taiwanese women
Early detection
Epidemiology

ABSTRACT

Background: Breast cancer has been the most frequently diagnosed cancer among women in Taiwan since 2003. While genetic variants play a significant role in the elevated risk of breast cancer, their implications have been less explored within Asian populations. Variant-based polygenic risk scores (PRS) have emerged as valuable tools for assessing the likelihood of developing breast cancer. In light of this, we attempted to establish a predictive breast cancer PRS tailored specifically for the Taiwanese population.

Methods: The cohort analyzed in this study comprised 28,443 control subjects and 1501 breast cancer cases. These individuals were sourced from the Taiwan Precision Medicine Initiative (TPMI) array and the breast cancer registry lists at Taichung Veterans General Hospital (TCVGH). Utilizing the breast cancer-associated Polygenic Score (PGS) Catalog, we employed logistic regression to identify the most effective PRS for predicting breast cancer risk. Subsequently, we subjected the cohort of 1501 breast cancer patients to further analysis to investigate potential heterogeneity in breast cancer risk.

Results: The Polygenic Score ID PGS000508 demonstrated a significant association with breast cancer risk in Taiwanese women with a 1.498-fold increase in cancer risk(OR = 1.498, 95 % CI(1.431–1.567, p=5.38×10^-68). Individuals in the highest quartile exhibited a substantially elevated risk compared to those in the lowest quartile, with an odds ratio (OR) of 3.11 (95 % CI: 2.70–3.59; p=1.15×10^-55). In a cohort of 1501 breast cancer cases stratified by PRS distribution, women in the highest quartile were diagnosed at a significantly younger age (p=0.003) compared to those in the lowest quartile. However, no significant differences were observed between PRS quartiles in relation to clinical stage (p=0.274), pathological stage (p=0.647), or tumor subtype distribution (p=0.244).

Conclusion: In our study, we pinpointed PGS000508 as a significant predictive factor for breast cancer risk in Taiwanese women. Furthermore, we found that a higher PGS000508 score was associated with younger age at the time of first diagnosis among the breast cancer cases examined.

1. Introduction

Breast cancer is the most prevalent malignancy and ranks second in terms of cancer-related fatalities among women worldwide [1]. Although the incidence and mortality rates of breast cancer in Asia are considerably lower than those in North America and Europe, recent reports have indicated a doubling in the incidence rate of breast cancer in Asia over the past few decades, thus establishing breast cancer as a significant public health concern [2,3]. In Taiwan, it is the most commonly diagnosed cancer among women, with incidence and

* Corresponding author.
** Corresponding author at: Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan.
E-mail addresses: gnlum@vghtc.gov.tw (C.Y. Lum), thsiao@vghtc.gov.tw (T.-H. Hsiao).

https://doi.org/10.1016/j.canep.2024.102701
Received 22 April 2024; Received in revised form 25 October 2024; Accepted 4 November 2024
Available online 19 December 2024

mortality rates on the rise [4]. However, Taiwan lags significantly behind Western countries in terms of breast cancer screening rates, with only 36 % of eligible individuals participating compared to 80 % in Western nations [5]. This discrepancy has resulted in delayed diagnoses and poorer survival rates. Given the constraints of breast cancer screening in Taiwan, the development of an appropriate screening strategy has become a top priority.

High-risk genetic factors, such as BRCA1 and BRCA2 mutations, are strongly associated with breast cancer risk, but account for less than 20 % of breast cancer cases [6]. In addition to these highly penetrant mutations, multiple common susceptibility variants individually confer modest risks. However, when combined into a polygenic risk score (PRS), these variants collectively exert a substantial influence [7,8]. Numerous studies have crafted variant-based PRSs for the purpose of stratifying lifetime breast cancer risk, with some of them finding their way into clinical practice [9–12]. However, most genome-wide association studies (GWAS) have primarily concentrated on European populations [13,14], leaving a gap in our understanding of genetic risk factors in specific Asian subpopulations leaving a gap in our understanding of genetic risk factors in specific Asian subpopulations. Recent research efforts have shifted towards crafting breast cancer PRSs tailored for Asian populations, which have demonstrated promising predictive capabilities [15,16]. For example, Ho WK et al. [16]developed and validated various PRSs designed specifically for East Asian women. Their observations revealed that PRSs generated by integrating data from both European and Asian ancestry GWAS datasets outperformed those relying solely on weights derived from single-ancestry GWAS data. Additionally, substantial disparities in PRS distributions were noted among different ethnic groups, highlighting the potential of these PRSs for predicting breast cancer risk in specific ancestral populations.

Building upon these recent advances, our study aims to address several crucial gaps in breast cancer risk prediction for the Taiwanese population. Despite the progress made in developing PRSs for East Asian populations, the genetic diversity within Asia necessitates population-specific studies. The Taiwanese population, with its unique genetic admixture of Han Chinese, indigenous Taiwanese, and various colonial influences [15], may harbor distinct genetic risk factors for breast cancer. Furthermore, the median age of breast cancer diagnosis in Taiwan is significantly younger than in Western populations, underscoring the need for tailored risk prediction models[17]. This comprehensive approach allow us to understand how genetic risk interacts with clinical presentation in the Taiwanese context. Through these objectives, we strive to provide crucial information for developing tailored, cost-effective screening and prevention strategies, potentially improving early detection of breast cancer in the Taiwanese population and addressing the current disparities in screening rates and survival outcomes.

2. Materials and methods

2.1. Study population

This retrospective cohort analysis, conducted in a hospital setting, encompassed 58,091 Taiwanese residents aged 20 and above. The research drew upon data from the Taiwan Precision Medicine Initiative (TPMI), a project spearheaded by Taiwan's Academia Sinica. The study period extended from June 2019 through May 2021. Initially, 8757 breast cancer patients were identified using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code 174. Their genetic profiles were linked to medical claims data from Taichung Veterans General Hospital (TCVGH).

To ensure data quality and consistency, we conducted a thorough review of medical records. The following inclusion criteria were applied:

1. Diagnosis confirmed through core needle biopsy performed at our hospital.
2. Subsequent treatment and follow-up completed at our hospital.
3. Pathological diagnosis of invasive breast cancer.
4. Complete reports of ER, PR, HER2, and Ki-67 status.
5. Genotyping results meeting our quality control standards.

Included patients need to meet all aforementioned criteria.After applying these stringent criteria, the final study cohort comprised 1501 eligible patients as shown in supplementary table 1. The control group comprised 28,443 women from the TPMI cohort at TCVGH who had not been diagnosed with breast cancer as presented in supplementary table 2.

This study was conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments. All participants provided informed consent for the use of their genetic and clinical data in ongoing research.

2.2. Genotyping and quality control

This study collected blood samples from all participants for DNA extraction. Genotyping was conducted using the Axiom Genome-Wide TWB 2.0 Array Plate (Affymetrix, Santa Clara, CA, USA). This array contains 714,431 SNPs specifically optimized for Taiwan's population [18], including approximately 300,000 known risk variants.Initial quality control and analysis were conducted using Affymetrix Power Tools software. Markers were excluded based on the following criteria: failure to meet Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-5}$), minor allele frequency below 0.05, or genotype missing rate exceeding 5 %. Subsequently, genotype imputation across autosomal chromosomes was conducted utilizing the Michigan Imputation Server, employing the 'minimac4' algorithm [19]. Prior to imputation, genotype data underwent strand alignment and were subsequently uploaded to the server. The imputation process leveraged the 1000 Genomes Phase 3 (Version 5) reference panel [20]. In our analysis, we incorporated all biallelic variants that met or exceeded an imputation quality threshold, defined by an INFO score of 0.3 or greater.

2.3. Polygenic risk score (PRS) selection process

Polygenic Risk Scores (PRSs) were computed using the 'score' function in PLINK v1.9 [21], a whole genome association analysis toolset.To identify the most effective [22](PRS) for predicting breast cancer risk in our Taiwanese cohort, we conducted a comprehensive screening using the Polygenic Score Catalog (PGS) (see Web Resources)., a publicly accessible database providing metadata for accurate PRS application and evaluation (Lambert et al., 2021). From this catalog, 104 PRSs was identified specifically reported to predict breast cancer risk. We employed logistic regression and standard deviation to assess the association of these PRSs with breast cancer in our cohort, retaining those demonstrating a significant association at $p < 10^{\wedge}-4$ for further analysis (as shown in Supplementary Table 3A and B). Subsequently, we subjected a cohort of 1501 breast cancer patients to additional analysis, investigating potential heterogeneity in breast cancer risk by examining factors such as age at diagnosis, tumor subtype, and clinical and pathological stages.

2.4. Statistical analysis

Logistic regression was utilized to assess the association between polygenic risk scores (PRSs) and breast cancer incidence. Both quantile-based Odds ratios (ORs) and Odds ratios per standard deviation increase in standardized PRSs were calculated with their corresponding 95 % confidence intervals (CIs).The study population was stratified into quartiles based on PRS distribution. Descriptive statistics, including

means, frequencies, and percentages, were employed to summarize clinical characteristics. Categorical variables were compared using chi-square tests. Kaplan-Meier analysis was performed to evaluate survival across breast cancer subtypes. A p-value < 0.05 was considered statistically significant.

3. Results

3.1. Comparative analysis of high-performing PRSs in breast cancer risk assessment

To identify the most predictive polygenic risk scores (PRSs) for breast cancer in our Taiwanese cohort, we screened 104 breast cancer-specific PRSs from the [22] Employing both quantile-based odds ratios and per standard deviation analyses, twelve PRSs emerged as significant predictors of breast cancer risk: PGS000508 [14], PGS000335 [23], PGS000507[14], PGS000511 [14], PGS002294 [24], PGS000007 [25], PGS000332 [23], PGS000512 [14], PGS000499 [14], PGS000500 [14], PGS000773 [26], and PGS000050[27] (Table 1A and Table 1B)

Logistic regression analysis demonstrated that PGS000508 exhibited superior predictive performance among all evaluated polygenic risk scores (PRSs). When stratifying the study population into quartiles based on PRS distribution, [14] PGS000508 showed the strongest association with breast cancer risk, with individuals in the highest quartile (Q4) demonstrating more than threefold increased risk compared to those in the lowest quartile (Q1) (OR = 3.11, 95 % CI: 2.70–3.59, p = 1.15×10^-55). For each standard deviation increase in PGS000508, the odds of developing breast cancer increased substantially (OR = 1.497, 95 % CI: 1.431–1.567, p = 1.13×10^-67).

While also showing strong predictive capability, the other two PRSs demonstrated slightly lower associations. PGS000335 [23] revealed that individuals in Q4 had a lower but still substantial risk compared to those in Q1 (OR = 3.05, 95 % CI: 2.65–3.52, p = 2.25×10^-55), with each standard deviation increase associated with elevated risk (OR = 1.497, 95 % CI: 1.431–1.567, p = 1.13×10^-67). Similarly, PGS000507 [14] showed marginally lower associations, with Q4 versus Q1 OR of 3.03 (95 % CI: 2.63–3.49, p = 2.42×10^-53) and per standard deviation OR of 1.493 (95 % CI: 1.426–1.562, p = 1.38×10^-66)

The superior discriminative ability of PGS000508 is visually demonstrated in Fig. 1(A), which illustrates the distinct distribution patterns between breast cancer cases and controls.

Development and characteristics of each of the top 10 PRSs, along with independent PGP IDs for reference with their respective statistics, are presented in Supplementary table 4

Table 1A
Top 10 PRSs: Quantile-Based Association with Breast Cancer Risk in Taiwanese Females (TCVGH TPMI).

PGS Catalog	Case	Control	OR Q4/Q1	P value Q4/Q1
	Mean (SD)	Mean (SD)	(95% CI)	
PGS000508	0.378 (0.214)	0.294 (0.212)	3.11 (2.70–3.59)	1.15E-55
PGS000335	0.171 (0.212)	0.088 (0.209)	3.05 (2.65–3.52)	2.25E-53
PGS000507	0.516 (0.214)	0.433 (0.212)	3.03 (2.63–3.49)	2.42E-53
PGS000511	0.630 (0.225)	0.548 (0.222)	2.84 (2.47–3.27)	5.16E-48
PGS002294	-0.444 (0.474)	-0.625 (0.475)	2.66 (2.32–3.06)	1.78E-44
PGS000007	-0.022 (0.390)	-0.156 (0.386)	2.65 (2.31–3.04)	1.07E-43
PGS000332	0.344 (0.205)	0.274 (0.203)	2.64 (2.30–3.03)	2.52E-43
PGS000512	2.340 (0.798)	2.054 (0.793)	2.63 (2.29–3.02)	7.28E-43
PGS000499	0.372 (0.252)	0.280 (0.249)	2.61 (2.28–3.00)	1.01E-42
PGS000500	0.208 (0.254)	0.118 (0.250)	2.62 (2.28–3.01)	1.13E-42

Table 1b
Top 10 PRSs: Association with Breast Cancer Risk in Taiwanese Females (TCVGH TPMI Database), Standard Deviation.

PGS Catalog	Case	Control	OR (95% CI)	P value
	Mean (SD)	Mean (SD)		
PGS000508	0.378 (0.214)	0.294 (0.212)	1.498 (1.431–1.567)	5.39E-68
PGS000335	0.171 (0.212)	0.088 (0.209)	1.497 (1.431–1.567)	1.13E-67
PGS000507	0.516 (0.214)	0.433 (0.212)	1.493 (1.426–1.562)	1.38E-66
PGS002294	-0.444 (0.474)	-0.625 (0.475)	1.486 (1.420–1.556)	1.28E-64
PGS000511	0.630 (0.225)	0.548 (0.222)	1.462 (1.396–1.531)	9.50E-59
PGS000512	2.340 (0.798)	2.054 (0.793)	1.454 (1.388–1.522)	2.11E-57
PGS000499	0.372 (0.252)	0.280 (0.249)	1.447 (1.383–1.514)	2.41E-57
PGS000773	0.352 (0.499)	0.186 (0.489)	1.439 (1.375–1.506)	8.04E-56
PGS000050	-0.112 (0.396)	-0.242 (0.390)	1.437 (1.373–1.503)	1.17E-55
PGS000500	0.208 (0.254)	0.118 (0.250)	1.439 (1.375–1.506)	1.60E-55

PRS polygenic risk score, OR odds ratio, TPMI Taiwan Precision Medicine Initiative
Q1, the first quartile distribution of the scores of PGS000508
Q4, the fourth quartile distribution of the scores of PGS000508
PRS polygenic risk score, OR odds ratio, TPMI Taiwan Precision Medicine Initiative

3.2. Age-dependent breast cancer risk assessment with PGS000508 quartiles

To further investigate the predictive power of PGS000508, we conducted an in-depth analysis categorizing women into groups based on the quartile distribution of this PRS. Women in the highest quartile exhibited a 1.95-fold increase in breast cancer risk compared to those in the lowest quartile, with a 95 % confidence interval ranging from 1.75 to 2.17 (p = 3.54 × 10^-34) (Table 2). The analysis of PGS000508 revealed significant differences in breast cancer incidence between the highest (Q4) and lowest (Q1) PRS quartiles, as illustrated in Fig. 1(B). The most pronounced divergence occurred between ages 45 and 50. By age 45, the cumulative incidence in the highest quartile (Q4) reached 30.6 %, compared to 18.2 % in the lowest quartile (Q1). This disparity widened by age 50, with Q4 showing a cumulative incidence of 50.7 % versus 34.9 % for Q1.

We further assessed the association between PGS000508 and breast cancer risk across different age categories as shown in Table 4. For the age groups of 30–40 and 40–50 years, the odds ratios for Q4 compared to Q1 were 2.36 and 2.40, respectively. This indicates that women in the highest quartile of PRS distribution in these age groups had more than 2fold higher risk of developing breast cancer compared to those in the lowest quartile

3.3. Subtype trends in breast cancer cases across PGS000508 quartiles

The demographic and disease characteristics of selected breast cancer cases are shown in Table 3. We divided the cohort into quartiles based on PGS000508, with 376 patients in Q1 and 375 patients each in Q2, Q3, and Q4. The mean ages for the 1st (Q1), 2nd (Q2), 3rd (Q3), and 4th (Q4) quartile distributions were 62 ± 10, 61 ± 10, 62 ± 11, and 59 ± 11 years, respectively. Notably, patients in Q4 had a significantly

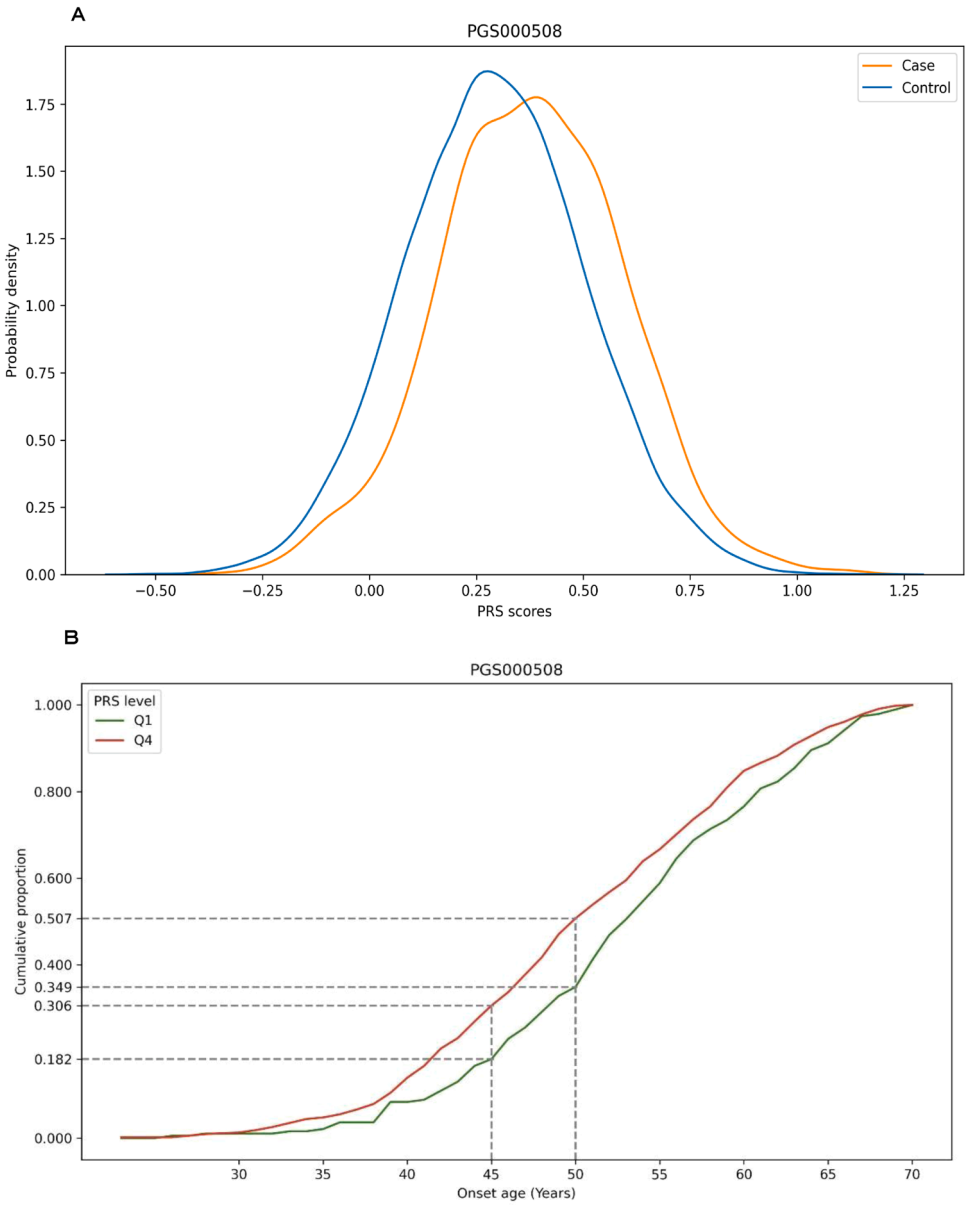

Fig. 1. (A)The probability density distribution of PGS000508 of the breast cancer and control groups, (B) The cumulative proportion by the age of onset in the highest quartile compared to the lowest quartile.

Table 2
Odds ratio(OR) of breast cancer in groups by quartile of the sores of PGS000508.

Risk score quartile	OR	95 %CI		P-value
1st(Q1)	1	reference		reference
2nd(Q2)	0.83	0.73	0.94	0.003
3rd(Q3)	1.05	0.93	1.18	0.426
4th(Q4)	1.95	1.75	2.17	<0.0001

younger age of breast cancer incidence ($p = 0.003$), suggesting that higher PRS scores may be associated with earlier onset of breast cancer.

Regarding disease characteristics, clinical stage 2 and pathological stage 1 were predominant among cases. In terms of subtype distribution, hormone receptor-positive breast cancer was the most common. However, it is worth noting that no significant differences were observed between PRS quartiles in relation to clinical stage ($p = 0.274$), pathological stage ($p = 0.647$), or tumor subtype distribution ($p = 0.244$). These findings indicate that while PGS000508 is strongly associated with breast cancer risk and age of onset, it may not be predictive of specific disease characteristics or subtypes.

3.4. Survival analysis

The survival rates of patients within the first quartile (Q1) and the fourth quartile (Q4) of PGS000508 exhibited no statistically significant differences, with p-value 0.16., as depicted in Fig. 2.

4. Discussion

The importance of genetic evaluation for early detection and prevention of breast cancer has grown considerably, particularly for young women and those with a strong family history of the disease. Our study validated various Polygenic Risk Scores (PRSs) for breast cancer in Taiwanese women, and it identified PGS000508 as the PRS most strongly predictive of breast cancer in this population. This finding is particularly noteworthy given the origin and composition of PGS000508. According to the PGS catalog[22], PGS000508, our

Table 3
Demographic and disease characteristics of breast cancer patients (cases) stratified by quartile of the PRS of PGS000508.

	Quartile of polygenic risk socre (PGS000508)								P value
	Q1 (N=376)		Q2 (N=375)		Q3 (N=375)		Q4 (N=375)		
mean age, SD, years	62 ± 10		61 ± 10		62 ± 11		59 ± 11		0.003
Clinical stage n,%									
stage 0	17	4.70 %	14	3.90 %	8	2.20 %	18	5 %	0.274
stage I	123	33.70 %	130	36 %	143	40.20 %	125	34.50 %	
stage II	173	47.40 %	169	46.80 %	159	44.70 %	162	44.80 %	
stage III	38	10.40 %	31	8.60 %	34	9.60 %	31	8.60 %	
stage IV	14	3.80 %	17	4.70 %	12	3.40 %	27	7.20 %	
Pathologycal stage n,%									
stage 0	2	0.60 %	11	3.30 %	10	2.90 %	8	2.40 %	0.647
stage I	167	49.40 %	156	47 %	165	48.70 %	163	49.10 %	
stage II	121	35.80 %	118	35.50 %	113	33.30 %	108	32.50 %	
stage III	39	11.50 %	37	11.10 %	41	12.10 %	38	11.40 %	
stage IV	9	2.70 %	10	3 %	10	2.90 %	15	4.50 %	
Subtype n, %									
hormone receptor positive	223	59.30 %	250	66.70 %	241	64.30 %	245	65.30 %	0.244
HER 2	108	28.70 %	82	21.90 %	101	26.90 %	93	24.80 %	
TNBC	45	12 %	43	11.50 %	33	8.80 %	37	9.90 %	

*A comparisons of categorical variables were analyzed using the Chi-square test
Q1, Q2, Q3 and Q4 represent the 1st, 2nd, 3rd and 4th quartile distribution of the scores of PGS000508
PRS polygenic risk score; HER2 human epidermal growth factor receptor 2; TNBC Triple-negative breast cancer

Table 4
Association between PGS000508 and breast cancer risk by age categories.

Age Group (years)	OR (Q4/Q1)	95 % CI		P-value
< 30	-	-	-	-
30–40	2.36	1.31	4.27	4.48×10^{-03}
40–50	2.40	1.85	3.10	3.17×10^{-11}
50–60	1.96	1.62	2.37	4.69×10^{-12}
60–70	1.84	1.51	2.24	2.08×10^{-09}
70–80	1.70	1.26	2.28	4.25×10^{-04}
> 80	2.75	1.45	5.21	1.87×10^{-03}

OR odds ratio
Q1, the first quartile distribution of the scores of PGS000508
Q4, the fourth quartile distribution of the scores of PGS000508

best-performing PRS, was developed using the PRS-CS approach based on European ancestry, comprising 213,946 breast cancer cases and 317, 922 controls with 1120,410 variants.

Our findings revealed that among the top 10 predictive PRSs, the majority were based on European ancestry. Interestingly, the only PRS derived entirely from East Asian ancestry (PGS000050) did not outperform purely or mixed European ancestry-based PRSs. This observation is consistent with the findings of Ho W-K et al. [28], who demonstrated that the 313-SNV PRS, specifically PGS000344 PRS287_BC, developed in a European ancestry cohort, was more predictive of breast cancer risk in the Asian population than PRSs derived from Asian data. Previous studies have also acknowledged limitations in developing powerful Asian-specific PRS due to the smaller sample sizes available for Asian genetic studies [29–31].

Based on the use of a quartile-based approach offers a clinically

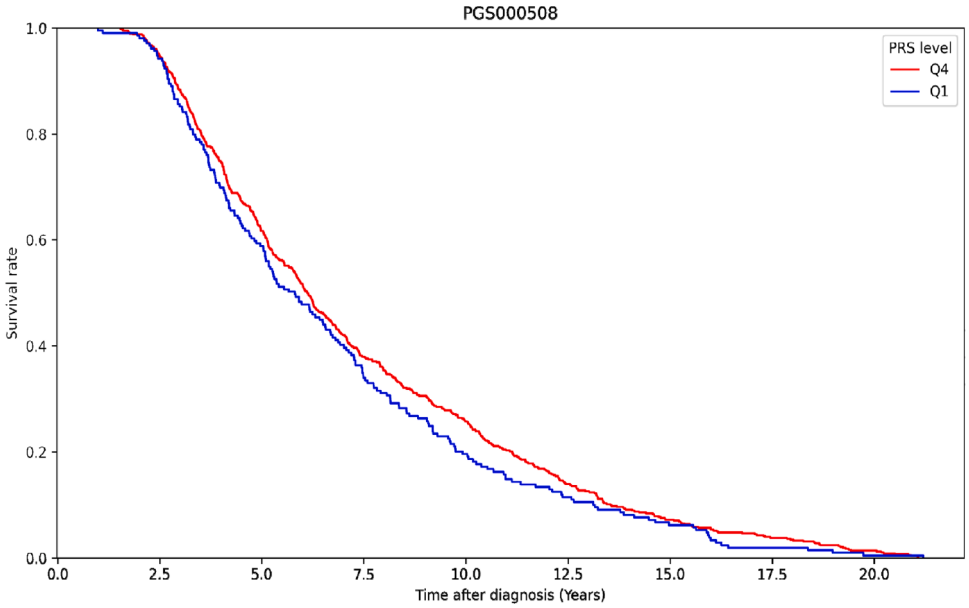

Fig. 2. The survival rates of patients within the first quartile (Q1) and the fourth quartile (Q4) of PGS000508.

intuitive framework for risk stratification, particularly in populations like the Taiwanese cohort, where genetic variant distributions may differ significantly from those observed in European populations. The decision to utilize quartile comparisons, rather than presenting odds ratios per standard deviation, was made to mitigate potential complexities in interpretation, especially in the context of non-linear relationships within the PRS distribution. Besides standard deviation, by focusing on quartile-based analysis, we highlighted risk extremes, which are often more pertinent for clinical decision-making when identifying high-risk individuals. This data-driven strategy enhances the practical application of PRS in risk prediction, providing a focused and actionable approach for clinical use. Our analysis of PGS000508 revealed striking differences in age at first breast cancer diagnosis between the highest (Q4) and lowest (Q1) PRS quartiles, particularly between ages 45 and 50. This pattern aligns with Taiwan's higher incidence of early-onset breast cancer [4]. The marked divergence in cumulative incidence between Q4 and Q1 in this critical age range highlights PGS000508's potential for identifying high-risk individuals in the Taiwanese population, suggesting its utility for targeted screening and early intervention strategies. These findings underscore the importance of age-specific risk assessment and the potential for personalized breast cancer prevention approaches in Taiwan.

In our study, women in the highest quartile, stratified by the distribution of the PRS of PGS000508, exhibited a 1.95-fold significantly increased risk of breast cancer compared to women in the lowest quartile as shown in Table 2. This finding demonstrates comparable or slightly lower predictive power when compared to previous PRSs established for Chinese women [12] and somewhat lower than the PRS developed for the Taiwanese population by Hsieh Y-C et al. [30]. Moreover, it is important to note that risk prediction models for breast cancer, which combine PRS with clinical risk factors, have been developed and shown to improve prediction in numerous studies [24,30, 32–34].These models have the potential to enhance the cost-effectiveness of screening programs and strike a better balance between the benefits and potential harms of screening[9,35,36]. This pattern diverges from typical PRS findings, which generally show a continuous risk gradient across quartiles. To investigate this phenomenon, we conducted a Phenome-Wide Association Study (PheWAS) exploring PRS associations with various clinical phenotypes [37,38]. The observed threshold effect may reflect population-specific genetic influences or limitations in applying European-derived PRSs to non-European cohorts. Our findings highlight the complexity of breast cancer risk assessment across diverse ethnic groups and underscore the need for population-specific PRS refinement.

Early-onset breast cancer, characterized by diagnoses occurring in women younger than 45 years old, is associated with a more aggressive tumor profile. It often presents with a high histological grade, an elevated proliferation rate, positive human epidermal growth factor receptor 2 status, negative hormone receptor expression, and a higher risk of local recurrence [39–43]. Kataoka A. et al. found that young age was an independent poor prognostic factor for 5-year disease-free survival, breast cancer-specific survival, and overall survival [44]. While the majority of breast cancer cases are diagnosed in women aged older than 45 years, a higher incidence of early-onset breast cancer has been observed in Eastern Asia [45]. Research conducted in China revealed that breast cancer was commonly diagnosed between the ages of 40 and 50, with a mean age of 48–49, which was more than 10 years younger than the reported average age in Western countries [46]. In Taiwan, the mean age of breast cancer diagnosis was 50.5 years in 1997 and 56.0 years in 2016; the median age at diagnosis was 48 and 55 years old, respectively [4]. In our study, we observed that individuals in quartile 4 (Q4) had a higher rate of early-onset breast cancer compared to those in quartile 1 (Q1). Furthermore, women in the highest quartile of polygenic risk score (PRS) had an approximately 2-fold higher risk of developing breast cancer between the ages of 30 and 50. These findings underscore the effective utility of PRS in predicting and identifying women at high risk for breast cancer, particularly in early-onset cases.

Previous research by Mavaddat et al., using data from women of European ancestry, reported that breast cancer Polygenic Risk Scores (PRSs) exhibited superior predictive performance for estrogen receptor (ER)-positive disease compared to ER-negative disease [10]. Furthermore, other studies have demonstrated that subtype-specific PRSs offer enhanced predictive accuracy for specific subtypes of breast cancer [16]. However, our analysis of PGS000508 in a Taiwanese cohort yielded different results. We found no statistically significant differences between PRS quartiles for clinical stage ($p = 0.274$), pathological stage ($p = 0.647$), or tumor subtype distribution ($p = 0.244$). This discrepancy with previous findings may be attributed to several factors, with sample size differences being a primary consideration. The substantially larger sample sizes in previous studies, particularly those focusing on European populations, may have provided greater statistical power to detect subtle differences in PRS performance across breast cancer subtypes. Our study's more limited sample size could have reduced our ability to identify similar subtype-specific associations, even if they exist in the Taiwanese population. This observation underscores the need for larger-scale studies in diverse populations to fully elucidate the performance of PRSs across different ethnic groups and breast cancer subtypes. It also highlights the importance of cautious interpretation when applying PRSs developed in one population to another, as genetic risk factors and their associations may vary across different ethnic backgrounds.

5. Conclusion

Our study, the first comprehensive validation of multiple Polygenic Risk Scores (PRSs) in a large Taiwanese cohort, identified PGS000508 as the most predictive PRS for breast cancer risk, demonstrating its transferability from European to Taiwanese populations. Unlike previous European studies, we found no significant differences in PRS performance across tumor subtypes or stages. These findings have important implications for developing tailored screening strategies in Taiwan, where breast cancer screening rates are lower than in Western countries. Future research should focus on larger-scale studies in diverse Asian populations and the integration of PRSs with clinical and lifestyle factors for more personalized risk assessment.

Web resources

Polygenic Score Catalog(PGS), http://www.pgscatalog.org

Funding

This study was funded by Academia Sinica (40–05-GMM, AS-GC-110-MD02 and 236e-1100202), and National Development Fund, Executive Yuan (NSTC 111–3114-Y-001–001).

CRediT authorship contribution statement

Chih-Chiang Hung: Supervision, Resources, Conceptualization. **Yi-Hsuan Lin:** Writing – original draft, Investigation. **I-Chen Tsai:** Resources. **Guan-Cheng Lin:** Visualization, Validation, Software, Methodology, Data curation. **Chih Yean Lum:** Writing – review & editing, Writing – original draft, Resources. **Tzu-Hung Hsiao:** Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

We thank all of the participants and investigators from the Taiwan Precision Medicine Initiative. The authors sincerely appreciate the assistance of the Center for Translational Medicine of Taichung Veterans

General Hospital. This study was funded by Academia Sinica (40–05-GMM, AS-GC-110-MD02 and 236e-1100202), and National Development Fund, Executive Yuan (NSTC 111–3114-Y-001–001).

Acknowledgments

The genotyping data were supported by Taiwan Precision Medicine Initiative (TPMI) program. We thank all of the participants and investigators from the Taiwan Precision Medicine Initiative. We sincerely appreciate the assistance of the Center for Translational Medicine of Taichung Veterans General Hospital.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.canep.2024.102701.

References

[1] E.T. Sedeta, B. Jobre, B. Avezbakiyev, Breast cancer: Global patterns of incidence, mortality, and trends, J. Clin. Oncol. 41 (16_) (2023) 10528, 10528.
[2] Y.X. Lim, et al., Breast Cancer in Asia: incidence, mortality, early detection, mammography programs, and risk-based screening initiatives, Cancers (Basel) 14 (17) (2022).
[3] F. Bray, et al., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 74 (3) (2024) 229–263.
[4] Y.C. Chen, et al., Forecast of a future leveling of the incidence trends of female breast cancer in Taiwan: an age-period-cohort analysis, Sci. Rep. 12 (1) (2022) 12481.
[5] D.R. Youlden, et al., The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality, Cancer Epidemiol. 36 (3) (2012) 237–248.
[6] N. Mavaddat, et al., Genetic susceptibility to breast cancer, Mol. Oncol. 4 (3) (2010) 174–191.
[7] N. Pashayan, et al., Polygenic susceptibility to prostate and breast cancer: implications for personalised screening, Br. J. Cancer 104 (10) (2011) 1656–1663.
[8] C. Wendt, S. Margolin, Identifying breast cancer susceptibility genes - a review of the genetic background in familial breast cancer, Acta Oncol. 58 (2) (2019) 135–146.
[9] N. Pashayan, et al., Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: a life-table model, JAMA Oncol. 4 (11) (2018) 1504–1510.
[10] N. Mavaddat, et al., Polygenic risk scores for prediction of breast cancer and breast cancer subtypes, Am. J. Hum. Genet 104 (1) (2019) 21–34.
[11] T. Yanes, et al., Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field, Breast Cancer Res 22 (1) (2020) 21.
[12 W.K. Ho, et al., Polygenic risk scores for prediction of breast cancer risk in Asian populations, Genet Med 24 (3) (2022) 586–600.
[13] S.L. Park, I. Cheng, C.A. Haiman, Genome-Wide Association Studies of Cancer in Diverse Populations, Cancer Epidemiol. Biomark. Prev. 27 (4) (2018) 405–417.
[14] L.G. Fritsche, et al., Cancer PRSweb: an online repository with polygenic risk scores for major cancer traits and their evaluation in two independent biobanks, Am. J. Hum. Genet 107 (5) (2020) 815–836.
[15] C.H. Chen, et al., Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan Biobank project, Hum. Mol. Genet 25 (24) (2016) 5321–5331.
[16] C. Hou, et al., Development and validation of polygenic risk scores for prediction of breast cancer and breast cancer subtypes in Chinese women, BMC Cancer 22 (1) (2022) 374.
[17] S.H. Cheng, et al., Unique features of breast cancer in Taiwan, Breast Cancer Res Treat. 63 (3) (2000) 213–223.
[18] C.-T. Fan, J.-C. Lin, C.-H. Lee, Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical Island, Pharmacogenomics 9 (2) (2008) 235–246.
[19] S. Das, et al., Next-generation genotype imputation service and methods, Nat. Genet. 48 (10) (2016) 1284–1287.
[20] A. Auton, et al., A global reference for human genetic variation, Nature 526 (7571) (2015) 68–74.
[21] A.T. Marees, et al., A tutorial on conducting genome-wide association studies: Quality control and statistical analysis, Int. J. Methods Psychiatr. Res. 27 (2) (2018) e1608.
[22] S.A. Lambert, et al., The polygenic score catalog as an open database for reproducibility and systematic evaluation, Nat. Genet 53 (4) (2021) 420–425.
[23] N. Mars, et al., Comprehensive inherited risk estimation for risk-based breast cancer screening in women, J. Clin. Oncol. 42 (13) (2024) 1477–1487.
[24] Y. Yang, et al., Incorporating polygenic risk scores and nongenetic risk factors for breast cancer risk prediction among Asian women, JAMA Netw. Open 5 (3) (2022) e2149030.
[25] N. Mavaddat, et al., Polygenic risk scores for prediction of breast cancer and breast cancer subtypes, Am. J. Hum. Genet. 104 (1) (2019) 21–34.
[26] Z. Du, et al., Evaluating polygenic risk scores for breast cancer in women of African ancestry, JNCI: J. Natl. Cancer Inst. 113 (9) (2021) 1168–1176.
[27] W. Wen, et al., Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry, Breast Cancer Res. 18 (1) (2016) 124.
[28] W.K. Ho, et al., European polygenic risk score for prediction of breast cancer shows similar performance in Asian women, Nat. Commun. 11 (1) (2020) 3833.
[29] W. Wen, et al., Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry, Breast Cancer Res 18 (1) (2016) 124.
[30] Y.C. Hsieh, et al., A polygenic risk score for breast cancer risk in a Taiwanese population, Breast Cancer Res Treat. 163 (1) (2017) 131–138.
[31] C.H.T. Chan, et al., Evaluation of three polygenic risk score models for the prediction of breast cancer risk in Singapore Chinese, Oncotarget 9 (16) (2018) 12796–12804.
[32] A. Lee, et al., BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors, Genet Med. 21 (8) (2019) 1708–1718.
[33] S.X. Li, et al., Prospective evaluation of the addition of polygenic risk scores to breast cancer risk models, JNCI Cancer Spectr. 5 (3) (2021).
[34] J.D. Brooks, et al., Validation of the International Breast Cancer Intervention Study (IBIS) model in the high risk ontario breast screening program: a retrospective cohort study, Genet Med 25 (6) (2023) 100820.
[35] G.L. Gierach, P.P. Choudhury, M. García-Closas, Toward risk-stratified breast cancer screening: considerations for changes in screening guidelines, JAMA Oncol. 6 (1) (2020) 31–33.
[36] P. Pal Choudhury, et al., Comparative validation of breast cancer risk prediction models and projections for future risk stratification, J. Natl. Cancer Inst. 112 (3) (2020) 278–285.
[37] Z. Ye, et al., Phenome-wide association studies (PheWASs) for functional variants, Eur. J. Hum. Genet 23 (4) (2015) 523–529.
[38] L. Wang, et al., Methodology in phenome-wide association studies: a systematic review, J. Med Genet 58 (11) (2021) 720–728.
[39] E. Yildirim, T. Dalgiç, U. Berberoğlu, Prognostic significance of young age in breast cancer, J. Surg. Oncol. 74 (4) (2000) 267–272.
[40] C.A. Gabriel, S.M. Domchek, Breast cancer in young women, Breast Cancer Res 12 (5) (2010) 212.
[41] B.M. Beadle, W.A. Woodward, T.A. Buchholz, The impact of age on outcome in early-stage breast cancer, Semin Radiat. Oncol. 21 (1) (2011) 26–34.
[42] D. Dai, et al., The prognostic impact of age in different molecular subtypes of breast cancer: a population-based study, PeerJ 7 (2019) e7252.
[43] A. Siddig, et al., The unique biology behind the early onset of breast cancer, Genes (Basel) 12 (3) (2021).
[44] A. Kataoka, et al., Young adult breast cancer patients have a poor prognosis independent of prognostic clinicopathological factors: a study from the Japanese Breast Cancer Registry, Breast Cancer Res Treat. 160 (1) (2016) 163–172.
[45] A. Hironaka-Mitsuhashi, et al., A tissue microRNA signature that predicts the prognosis of breast cancer in young women, PLoS One 12 (11) (2017) e0187638.
[46] T. Li, C. Mello-Thoms, P.C. Brennan, Descriptive epidemiology of breast cancer in China: incidence, mortality, survival and prevalence, Breast Cancer Res Treat. 159 (3) (2016) 395–406.