Check for updates

Received: 25 January 2024 | Revised: 22 March 2024 | Accepted: 3 April 2024

DOI: 10.1002/cncr.35337

ORIGINAL ARTICLE

Population-specific validation and comparison of the performance of 77- and 313-variant polygenic risk scores for breast cancer risk prediction

Milena Hovhannisyan MSc[1] | Petra Zemankova PhD[1,2] | Petr Nehasil MSE[1,2,3] | Katerina Matejkova MSc[1,4] | Marianna Borecka PhD[1] | Marta Cerna MSE[1] | Tatana Dolezalova MSc[1] | Lenka Dvorakova PhD[3] | Lenka Foretova MD, PhD[5] | Klara Horackova MSc[1] | Sandra Jelinkova MSc[1] | Pavel Just MSc[1] | Marta Kalousova MD, PhD[1] | Jan Kral PhD[1,6] | Eva Machackova PhD[5] | Barbora Nemcova MSc[1] | Marketa Safarikova PhD[1] | Drahomira Springer MSE, PhD[1] | Barbora Stastna MSc[1,7] | Spiros Tavandzis MSc[8] | Michal Vocka MD, PhD[9] | Tomas Zima MD, PhD[1] | Jana Soukupova PhD[1] | Petra Kleiblova MD, PhD[1,10] | Corinna Ernst PhD[11] | Zdenek Kleibl MD, PhD[1,2] | Marketa Janatova PhD[1]

Correspondence
Marketa Janatova, Institute of Medical Biochemistry and Laboratory Diagnostics, First Faculty of Medicine, Charles University and General University Hospital in Prague, 121 08 Prague, Czech Republic.
Email: mjana@lf1.cuni.cz

Funding information
Ministry of Education, Youth and Sports of the Czech Republic, Grant/Award Number: LX22NPO5102; Ministry of Health of the Czech Republic, Grant/Award Numbers: DRO-VFN-64165, NU20-09-00355; Grant Agency, Charles University, Grant/Award Numbers: COOPERATIO, SVV260631, UNCE/24/MED/022

Abstract

Background: The polygenic risk score (PRS) allows the quantification of the polygenic effect of many low-penetrance alleles on the risk of breast cancer (BC). This study aimed to evaluate the performance of two sets comprising 77 or 313 low-penetrance loci (PRS77 and PRS313) in patients with BC in the Czech population.

Methods: In a retrospective case-control study, variants were genotyped from both the PRS77 and PRS313 sets in 1329 patients with BC and 1324 noncancer controls, all women without germline pathogenic variants in BC predisposition genes. Odds ratios (ORs) were calculated according to the categorical PRS in individual deciles. Weighted Cox regression analysis was used to estimate the hazard ratio (HR) per standard deviation (SD) increase in PRS.

Results: The distributions of standardized PRSs in patients and controls were significantly different ($p < 2.2 \times 10^{-16}$) with both sets. PRS313 outperformed PRS77 in categorical and continuous PRS analyses. For patients in the highest 2.5% of PRS313, the risk reached an OR of 3.05 (95% CI, 1.66–5.89; $p = 1.76 \times 10^{-4}$). The continuous risk was estimated as an $HR_{per\ SD}$ of 1.64 (95% CI, 1.49–1.81; $p < 2.0 \times 10^{-16}$), which resulted in an absolute risk of 21.03% at age 80 years for

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *Cancer* published by Wiley Periodicals LLC on behalf of American Cancer Society.

individuals in the 95th percentile of PRS313. Discordant categorization into PRS deciles was observed in 248 individuals (9.3%).

Conclusions: Both PRS77 and PRS313 are able to stratify individuals according to their BC risk in the Czech population. PRS313 shows better discriminatory ability. The results support the potential clinical utility of using PRS313 in individualized BC risk prediction.

KEYWORDS
breast cancer, germline genetic testing, polygenic risk score, PRS77, PRS313

INTRODUCTION

Breast cancer (BC) is the most prevalent oncologic disease, with a cumulative lifetime risk exceeding 10% in females in developed countries.[1] The risk of BC depends on various environmental and biological factors, among which family BC history is one of the strongest predictors.[2]

Although sporadic cases with no family history of BC comprise a majority of patients, 5%–10% of cases of BC can be explained by the presence of germline pathogenic variants (PVs) in high and moderate BC predisposition genes.[3] Compared to patients with sporadic BC, patients with hereditary BC are characterized by a different spectrum of BC subtypes: tumor development at an earlier age, increased risk of second primary cancers, and increased prevalence of breast, ovarian, and other related cancer types in relatives.[4] However, more than 80% of patients with a positive BC family history are not carriers of a PV in established cancer predisposition genes.[3]

Recent studies have suggested that part of the missing heritability in cases of BC can be attributed to the interplay among low-penetrance variants, usually single-nucleotide polymorphisms (SNPs).[5,6] Although each such SNP can modify BC risk only very slightly (but statistically significantly), their combined predictive power, expressed as a polygenic risk score (PRS), may lead to exceeding relative risk (RR) thresholds for moderate (RR, >2) or even high (RR, >4) BC risk.[7–9]

The PRS expresses BC risk at a level of polygenic genetic background, which summarizes the overall contributions of individual SNPs multiplied by the corresponding allelic dosages.[9–11] PRS values are assumed to be distributed normally among a correspondingly large group of individuals.[12,13] Because allelic frequencies of SNPs considered for the PRS differ in various populations, the raw PRS value must be compared to the mean PRS in the corresponding population.

The PRS may improve the stratification of BC risk in personalized medicine in assessing the proper management and timing of BC prevention.[14] To date, several SNP sets have been created on the basis of data from the Collaborative Oncological Gene-Environment Study and OncoArray.[10,15–17] Sets of 77 SNPs[10] and 313 SNPs[16] have been developed from a sample of women of European descent[8] and represent the gold standard in PRS analysis in female patients with BC. Nevertheless, the proband sample on which the sets were tested was heavily biased toward women of Western/Northern European descent. Before the application of these sets for BC risk prediction in other populations, the SNP sets need to be validated and standardized accordingly.

The focus of this study was to calculate and compare the predictive performance of the PRS from the sets of 77 SNPs[10] and 313 SNPs[16] and to test each set on samples from 1698 Czech female patients with BC and 1459 noncancer female control individuals. Carriers of PVs in known BC predisposition genes were excluded from both groups before the analysis.

MATERIALS AND METHODS

Study participants

The patient cohort consisted of 1698 female patients with BC (aged 20–80 years), including 481 patients with BC indicated for germline genetic testing on the basis of national testing guidelines[18] and 1217 nonindicated patients with BC who underwent germline genetic testing within clinical research projects. All patients with BC were enrolled at the General University Hospital (GUH) in Prague between 1997 and 2023. The control group consisted of 1459 noncancer individuals, including the following subgroups: (1) 405 healthy female volunteers (aged 60–80 years) with no family cancer history were enrolled at the GUH in Prague, Masaryk Memorial Cancer Institute (MMCI), and AGEL Laboratories between 2019 and 2023; and (2) 1054 women without cancer unselected for family cancer history (aged 20–80 years, enrolled between 2015 and 2023) included 97 female volunteers from the MMCI and 957 females enrolled at the GUH in Prague from patients tested because of various noncancer conditions (mainly cardiomyopathy and diabetes mellitus). All study participants were Czechs (Central Europeans). All participants signed informed consent forms for study participation, which was approved by the Ethics Committee of the GUH in Prague. All samples were anonymized before testing. Carriers of PVs in BC predisposition genes were excluded from PRS analysis.

Next-generation sequencing panel analysis

Genomic DNA was isolated from peripheral venous blood. All DNA samples were tested by targeted custom-designed next-generation

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sequencing (NGS) panels (KAPA HyperCap; Roche, Basel, Switzerland). The CZECANCA panel (Czech Cancer Panel for Clinical Application) has been used to exclude carriers of PVs in established BC predisposition genes, and the PRSMAN panel was designed to genotype variants for the PRS analysis.

The CZECANCA panel targeted 226 known and candidate cancer predisposition genes and was described in detail previously.[19–21]

The PRSMAN panel was designed to target 883 PRS-associated variants in breast, ovarian, and other cancer types that were published by July 2020 (Table S1).[8,10,15–17,22–35] The Roche probe design algorithm excluded 39 SNPs from the PRSMAN panel before its production. The final 844 targeted SNPs are listed in Table S2.

The NGS libraries were prepared with KAPA HyperPlus/EvoPlus Library Preparation Kits (Roche) according to the manufacturer with minor modifications described previously.[19,21,36] The target regions were captured with pooled CZECANCA and PRSMAN panels in a 2:1 ratio. Sequencing was performed with NextSeq 500 instruments (Illumina, San Diego, California).

Description of 77- and 313-SNP sets

We validated sets of 77 and 313 SNPs (henceforth referred to as PRS77 and PRS313) developed by Mavaddat et al. in 2015 and 2019.[10,16] Although all PRS77 SNPs were successfully genotyped, 48 SNPs from the PRS313 set were excluded from further analysis: 22 SNPs because of panel design restrictions (two of which were substituted by other SNPs in linkage disequilibrium covered in the PRSMAN panel; Table S3), 24 SNPs because of their localization in repetitive sequences, and two SNPs that did not match the allele specification in the PRS313 definition. Finally, 267 out of the 313 SNPs were genotyped by the PRSMAN panel. The overlap between the sets consisted of 15 SNPs.

Bioinformatics analysis

Raw NGS data were processed by an in-house bioinformatics pipeline as described previously.[19] The sequencing data in the FASTQ format were generated from NextSeq with the Illumina BaseSpace Sequence Hub (https://support.illumina.com/sequencing/sequencing_software/basespace.html). The FASTQ files were then mapped to the hg19 reference genome with NovoAlign version 2.08.03 (http://www.novocraft.com/products/novoalign). GATK version 3.3 HaplotypeCaller (https://github.com/broadinstitute/gatk-docs) was used for PRS SNP calling.

Via in-house scripts a call rate filter was applied, which excluded samples or SNPs with a call rate of <95% due to observed read coverages of <10× from further analysis. The set of SNPs was tested for Hardy–Weinberg equilibrium (HWE). Identity-by-state analysis was performed in PLINK version 1.90b6.26 (https://www.cog-genomics.org/plink) to prevent duplicated samples. Missing genotypes were substituted by the corresponding observed allele frequencies across patients and controls as described previously.[37] Created VCF files were used for subsequent PRS calculations.

PRS calculation

An individual's raw PRS was calculated as follows:

$$\mathrm{PRS} = \sum_{i=1}^{N} \log\mathrm{OR_i} \times \text{number of effect alleles},$$

namely the sum of the corresponding log OR and the number of observed effect alleles, which took the values 0, 1, or 2 (or dosage in case of imputation), per SNP. PRS calculation was performed in two steps: the first considered data from controls only to determine the mean and standard deviation (SD), and the second was performed to retrieve normalized PRS values both from control and patient data. PRS values were calculated separately for the PRS77 and PRS313 SNP sets, with the effect sizes of individual SNPs as published in the Polygenic Score (PGS) Catalog (PGS000001 for PRS77; PGS000004 for PRS313).[38]

Statistical analysis

Statistical analyses were performed with R version 4.2.0 (https://cran.r-project.org/). Differences between the standardized PRS values of patients and controls were assessed via the two-sided *t*-test, with *p* values of less than .05 considered statistically significant.

We compared the numbers of patients and controls in distinct deciles of the PRS to the reference middle quintile (5th and 6th deciles) to assess the OR in the respective groups. Statistical significance was calculated by the Fisher exact test in R.

The association of the PRS with BC development was determined by weighted Cox regression analysis with age at diagnosis as the time-dependent variable. To account for the oversampling of cases of BC in our retrospective study sample, real-world incidence rates in the Czech Republic in the years 2012–2022 collected by the Czech National Cancer Registry[39] were mimicked with the weighted cohort approach proposed by Antoniou et al.[40] Absolute age-specific risks of BC at distinct standardized PRS percentiles were estimated as a proportional hazards assumption in age-stratified Cox regression models as described by Zhang et al.[41] Age-specific cumulative BC risks were calculated as described previously.[27,42]

RESULTS

Selection of samples for PRS analysis and SNP validation

The initial sample set included 1698 patients with BC (aged 20–80 years) and 1459 controls (aged 20–80 years). Of these, 311

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

patients with BC (18.3%) and 55 control individuals (3.8%) harbored any germline PV in the established BC predisposition genes (i.e., *BRCA1*, *BRCA2*, *CHEK2*, *ATM*, *BARD1*, *BRIP1*, *MLH1*, *MSH6*, *NBN*, *NF1*, *PALB2*, *PMS2*, *PTEN*, *RAD51C*, *RAD51D*, and *TP53*; data not shown) and were excluded from subsequent PRS analyses. Thus, data from 1387 PV-negative patients with BC and 1404 PV-negative controls were considered for further analyses. The call rate filter excluded 43 patients and 0 controls for the PRS77 data set and 57 patients and 80 controls for the PRS313 data set. However, for the sake of direct comparability, only probands who passed the call rate filters for *both* the PRS77 and PRS313 data sets were included in further analyses, which yielded final data sets of 1329 patients with BC and 1324 controls (mean age, 56 and 54 years, respectively; $p = .15$ in the two-sided Welch *t*-test; Figure S1). The noncancer controls included two subsets with and without selection for family cancer history. The statistical analysis of raw PRS values in these subsets for PRS77 and PRS313 showed no statistically significant differences ($p = .06$ for PRS77; $p = .16$ for PRS313), and hence both control subsets were combined for further analyses.

All SNPs passed the call rate filter with a threshold of >95% for PRS77. Regarding the 267 SNPs genotyped within the PRS313 set, eight SNPs did not pass the quality filters for variant calling recommended by GATK, the SNP-specific call rate filter excluded a further 16 SNPs, and two SNPs did not pass HWE control, which resulted in a final set of 241 loci for further PRS313 analyses.

Association between PRS and BC risk

The distributions of standardized PRSs in patient and control samples are shown in Figure 1. Standardized PRS77 values ranged from −2.56 to 4.35 (mean, 0.34) in patient samples and −2.80 to 3.75 (mean, 0) in control samples. Standardized PRS313 values ranged from −2.85 to 3.72 (mean, 0.48) in patient samples and −3.34 to 3.13 (mean, 0) in control samples. For both SNP sets, the mean PRS of patient samples was statistically significantly increased compared to the mean PRS of control samples (PRS77, $p < 2.2 \times 10^{-16}$; PRS313, $p < 2.2 \times 10^{-16}$).

We calculated ORs for patients and control individuals binned by decile of the standardized PRS (Table 1; Table S4), with the middle quintile (5th and 6th deciles) as the reference. Although BC risk positively correlated with the accumulation of risk alleles in both data sets, PRS313 displayed more stably increasing risks and discriminated the risks better between the percentile categories (Figure 2; Table S4). Moreover, in the extremes of the distribution, the difference in calculated ORs was more pronounced in the PRS313 than in the PRS77 data set. For patients in the highest 2.5% of PRS313, the BC risk was calculated as an OR of 3.05 (95% CI, 1.66–5.89; $p = 1.76 \times 10^{-4}$). No such increase was observed in the highest 2.5% of PRS77.

To calculate hazard ratios (HRs) per SD for both SNP sets, we analyzed the association of the standardized PRS with BC risk via Cox regression. Consistent with the results above, the association with BC risk per SD was lower for PRS77 ($HR_{per\ SD}$, 1.40; 95% CI, 1.28–1.53; $p = 3.9 \times 10^{-15}$) than for PRS313 ($HR_{per\ SD}$, 1.64; 95% CI, 1.49–1.81; $p < 2.0 \times 10^{-16}$).

Age-specific HR estimates were used to calculate the absolute cumulative BC risk per PRS percentile with BC incidence data from the Czech National Cancer Registry.

PRS313 outperformed PRS77 because it assessed the risk of disease with significantly better consistency within a wider range of age of disease onset categories. This was also reflected in the global *p* values obtained (PRS77, $p = .15$; PRS313, $p = .01$; Table S5).

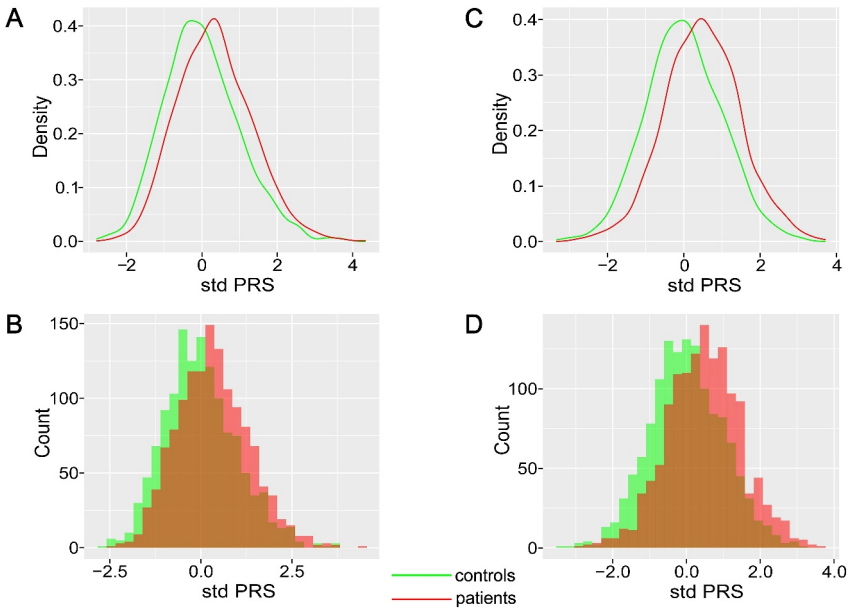

FIGURE 1 Empirical distributions of the standardized PRS of control (green) and patient samples (red). Both PRS sets PRS77 (A,B) and PRS313 (C,D) present statistically significant differences in mean standardized PRS values between the groups. PRS indicates polygenic risk score; std, standardized.

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Association between BC risk and deciles of standardized PRS77 and PRS313.

	PRS77		PRS313	
Decile	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***
1st	0.51 (0.37–0.70)	1.91×10^{-5}	0.38 (0.28–0.53)	2.59×10^{-9}
2nd	0.72 (0.53–0.98)	.03	0.55 (0.40–0.75)	1.07×10^{-4}
3rd	0.68 (0.50–0.92)	.01	0.66 (0.49–0.90)	.01
4th	0.89 (0.65–1.21)	.48	0.88 (0.65–1.20)	.45
5th and 6th	1	Reference	1	Reference
7th	1.34 (0.99–1.83)	.06	1.23 (0.91–1.68)	.19
8th	1.17 (0.86–1.59)	.34	1.46 (1.07–2.00)	.02
9th	1.59 (1.17–2.18)	2.97×10^{-3}	1.61 (1.18–2.20)	2.48×10^{-3}
10th	1.55 (1.14–2.12)	4.93×10^{-3}	1.87 (1.37–2.58)	6.81×10^{-5}

Abbreviations: BC, breast cancer; OR, odds ratio; PRS, polygenic risk score.

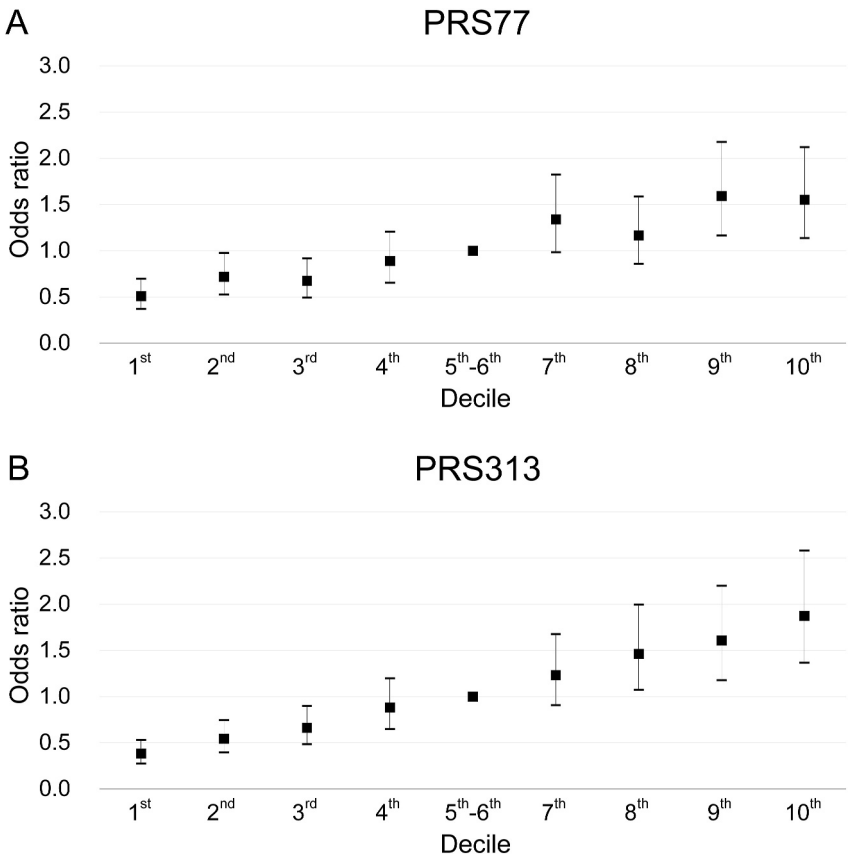

FIGURE 2 Association between PRS77 (A) and PRS313 (B) deciles and the occurrence of primary breast cancer. Odds ratios are calculated for each decile of the PRS relative to the reference (5th and 6th deciles) of the PRS. Ranges correspond to 95% CI. PRS indicates polygenic risk score.

PRS313 better discriminates the risks between the 5th and 95th percentiles (Table 2; Figure 3), which improved cancer risk estimates over PRS77 in both the age categories of 50 years (1.46 vs. 4.90 for PRS77; 1.23 vs. 4.85 for PRS313) and 80 years (5.78 vs. 18.93 for PRS77; 4.54 vs. 21.03 for PRS313). In addition, the estimated risk for individuals at the 95th percentile exceeds 20% at the age of 80 years only for PRS313 but not for PRS77.

In addition, we evaluated the results of both PRS77 and PRS313 on an individual level. We compared decile categorization for all individuals successfully genotyped in both sets ($n = 2653$). Surprisingly,

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Absolute risks for the 5th, 10th, 50th, 90th, and 95th percentiles of PRS77 and PRS313 by the ages of 50 and 80 years.

	PRS77 BC risk		PRS313 BC risk	
PRS percentile	Age 50 years, %	Age 80 years, %	Age 50 years, %	Age 80 years, %
5th	1.46	5.78	1.23	4.54
10th	1.57	6.48	1.41	5.36
50th	2.22	9.84	2.35	9.73
90th	3.84	15.81	4.08	17.67
95th	4.90	18.93	4.85	21.03

Abbreviations: BC, breast cancer; PRS, polygenic risk score.

both sets identically categorized only 502 individuals (18.9%), and the categorization over five deciles differed for 248 individuals (9.3%) (Figure 4).

DISCUSSION

Prevention is a key strategy to lower BC burden, as has been demonstrated by the reduction of BC-related mortality since the introduction of mammography for population screening programs.[43,44] Although the value of mammography is unquestionable, further BC reduction requires personalized approaches able to categorize females with different nonmodifiable risks. Various risk assessment models, including the Gail/Breast Cancer Risk Assessment Tool, Claus, Tyrer-Cuzick, and BOADICEA (Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm), based on the evaluation of family and personal gynecological/cancer history and clinical characteristics, have been used.[45] Nevertheless, they might not predict risk accurately for women who do not fit into the risk groups for whom they were developed.

Polygenic inheritance has been identified as an important genetic factor modifying BC risk. To date, the two most commonly used sets of variants, PRS77 and PRS313,[10,16] have been validated in various populations.[26,31,37,46–52] Because the effect of SNPs on BC risk may vary among different ethnicities, in this study, we tested their predictive performance in a population of Czech women, which has not been analyzed to date. Recently, Yiangou et al. reported a comparison of PRS313 performance in various European populations and found significant differences between means (−0.12 to 0.25), which can influence the estimation of individual risk and suggest using country-specific PRS distributions to calibrate risk categories.[53]

The comparison of our study in the context of previous work is complicated because of the great heterogeneity in the reporting of the resulting risk estimates (OR vs. HR), given the diverse bioinformatical methods used and the varying numbers of participants.

PRS77 was constructed in 2015 on the basis of samples from 33,673 female patients with BC and 33,381 control individuals collected by the Breast Cancer Association Consortium (BCAC).[8,22] The study identified that PRS77 was associated with BC risk with an $OR_{per\ SD}$ of 1.55.[10] Similar results were observed in a subsequent

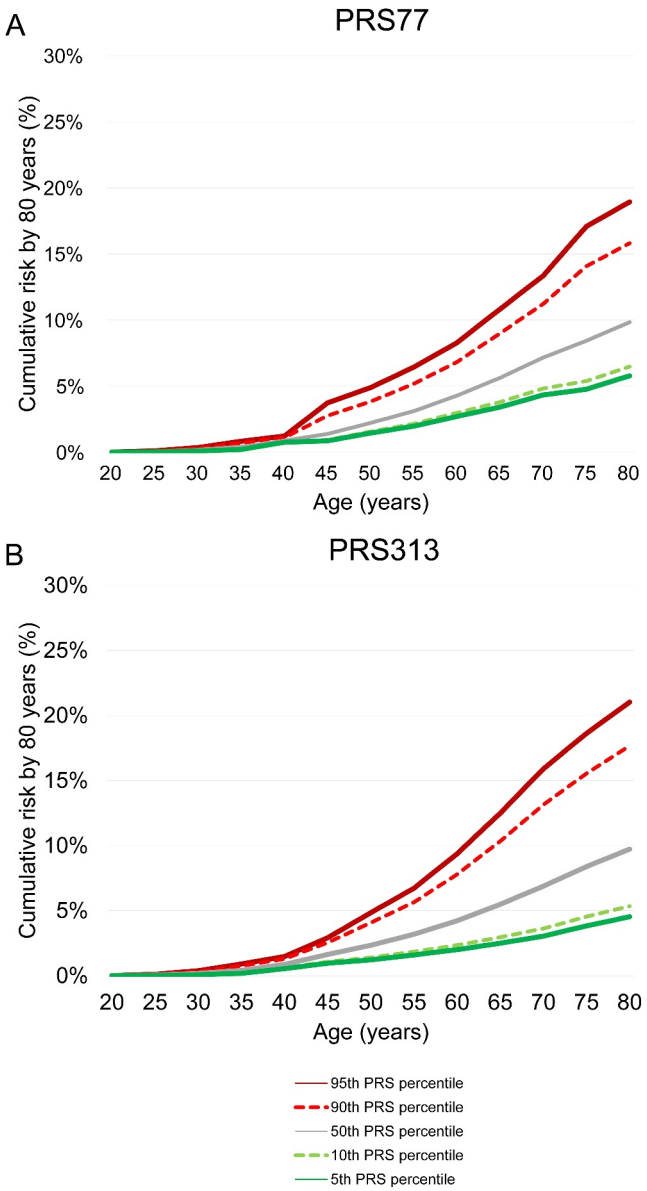

FIGURE 3 Predicted cumulative breast cancer risks according to the percentile of the PRS via sets PRS77 (A) and PRS313 (B). PRS indicates polygenic risk score.

Australian study, which was performed on a much smaller sample set (750 patients with BC and 405 controls), with an $OR_{per\ SD}$ of 1.46.[46] Our results are consistent with these findings ($HR_{per\ SD}$, 1.40).

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

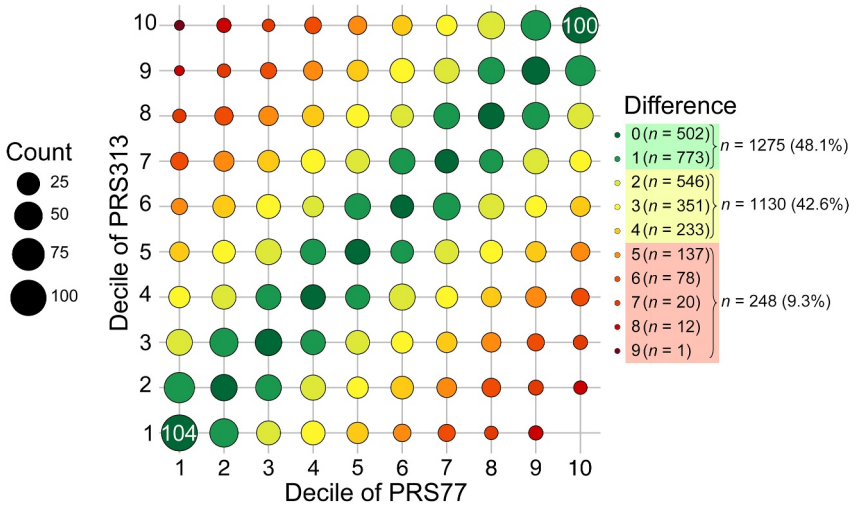

FIGURE 4 Visualization of the discordance in the stratification of PRS values per proband between PRS77 and PRS313. The size of the dots corresponds to the number of probands per category. PRS indicates polygenic risk score.

Regarding PRS77 by percentile, we observed that women in the top PRS decile (90%–100%) were at higher risk (OR, 1.55) when compared to women in the middle quintile (5th and 6th deciles), yet no substantial risk increase was observed for patients in the top 2.5% of PRS77 (OR, 1.55). However, in the original BCAC study, the risk of BC was three times higher for women in the 99th percentile than for women in the middle quintile (OR, 3.36).[10]

PRS313 was constructed in 2019 on the basis of data from 94,075 patients with BC and 75,017 control individuals from expanded BCAC studies.[8,22,23] It states the BC risk as an $OR_{per\ SD}$ of 1.65 in the perspective set and 1.61 in the validation set. In our study, the BC risk for the continuous PRS313 set was assessed as an $HR_{per\ SD}$ of 1.64. Thus, our results correspond to the initial study from 2019 as well as the subsequent Dutch study by Lakeman et al. ($OR_{per\ SD}$, 1.97).[54]

Regarding PRS313 by percentile, we found that women within the PRS top decile (90%–100%) have a higher risk (OR, 1.87) than those in the middle quintile (5th and 6th deciles). Moreover, the BC risk was increased (OR, 3.05) for patients in the highest 2.5% of the PRS. In the original study, the risk was four times higher for women in the 99th percentile in comparison to the middle quintile. A large UK cohort study by Jia et al. described the BC risk for patients in the 5th quintile as double when compared to the middle quintile.[31]

These results suggest that both PRS77 and PRS313 allow the identification of increased BC risk and discriminate between noncancer individuals and patients with BC in the Czech population. However, in our study, PRS313 had a higher discrimination power regarding both when applied in distinct percentile categories and continuously. Two previous studies by Tasa et al. and Jiao et al. also compared both sets in other populations,[48,49] and both studies reported better performance of PRS313 as well.

Mavaddat et al. described that PRSs considering larger numbers of variants discriminate better between patients with BC and noncancer controls.[55] Besides the number of SNPs, the PRS sets differ in the effect sizes of the SNPs included. Therefore, the overlap of only 15 SNPs between PRS77 and PRS313 may lead to a discordant classification of individuals analyzed via different SNP sets, as demonstrated in our results. Such discrepancies could lead to discordant reports for patients with potentially negative effects on their clinical management, which indicates a need for standardization of PRS analysis in clinical settings. The predictive value of PRS77 can be overestimated as a result of using the same patient cohort for development and evaluation[56] and biased because of the design using predominantly later onset patients.[46] The large differences between PRS77 and PRS313 shown in our and other studies call for careful validation of both sets under precise criteria in well-defined case-control data from various populations. This will require the establishment of consensus guidelines for statistical analysis and risk calculation.

Regarding absolute cumulative BC risk, the risk reaches 21.02% for individuals in the highest 5% of PRS313. Similar results were described in the initial study by Mavaddat et al. (19% risk for the 90th–95th PRS percentile; 24% risk for the 95th–99th PRS percentile).[16] An overall risk above 20% exceeds a general threshold for clinical management of individuals at BC risk used for carriers of germline PVs in moderate-penetrance genes in the Czech Republic.[18]

Our data indicate that individuals in the lowest 5% of PRS313 reach the BC risk of the average population at age 55 years. On the other hand, individuals in 95% of PRS313 reach the BC risk of the average population at age 40 years. This is consistent with previously published data suggesting deferral of mammogram screening for women at low risk and earlier screening initiation and shorter time intervals between screening visits for women at high risk.[57,58] However, further studies are needed to evaluate the impact of the PRS on BC onset and the related personalized start of screening.

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Currently, the utility of PRS implementation in clinical practice is still a matter of ongoing debate regarding the following issues. The discriminatory power of the PRS should be sufficient to demonstrate its ability to accurately identify individuals at risk of developing the disease. Its performance should be compared with existing predictive methods; even a score with high discriminatory power may be redundant if better predictors are already in clinical use. Appropriate preventive measures must also be available. Well-controlled clinical trials will be needed to fully address these requirements.

A limitation of our study was the lack of information on BC risk factors (i.e., family history) or BC subtypes in the nonindicated group of patients with BC. In order to increase the precision of the risk prediction, the interaction between the PRS and clinical risk factors ought to be considered in future analyses. Further development of a population-specific SNP set validated with extended cohorts of BC cases and controls may improve their discriminatory power.

In conclusion, in this study, we have assessed that the previously designed PRS313 can reliably discriminate groups of patients with BC and noncancer individuals from the Czech population, which stratifies women according to their BC risk. On the other hand, comparison of the two most frequently validated BC sets, PRS77 and PRS313, demonstrated large differences in an individual's classification, which warrants an urgent need for standardization of PRS analysis in routine clinical settings. Moreover, the PRS cannot act as a stand-alone test for BC risk prediction. For the best performance, risk estimates must be combined with the assessment of other genetic (analysis of moderate/high-penetrance cancer predisposition genes) and nongenetic factors (lifestyle, age of onset, and family history). Robust prospective clinical trials that include all risk factors, as well as the PRS, are needed in the future.

AUTHOR CONTRIBUTIONS

Milena Hovhannisyan: Writing–original draft; writing–review and editing; data curation; investigation; validation; and visualization. **Petra Zemankova**: Data curation; formal analysis; investigation; software; visualization; and writing–review and editing. **Petr Nehasil**: Data curation; investigation; software; and writing–review and editing. **Katerina Matejkova**: Visualization; formal analysis; investigation; software; validation; and writing–review and editing. **Marianna Borecka**: Investigation; validation; and writing–review and editing. **Marta Cerna**: Investigation; writing–review and editing; and validation. **Tatana Dolezalova**: Validation and writing–review and editing. **Lenka Dvorakova**: Resources and writing–review and editing. **Lenka Foretova**: Resources and writing–review and editing. **Klara Horackova**: Investigation; writing–review and editing; and validation. **Sandra Jelinkova**: Investigation; validation; and writing–review and editing. **Pavel Just**: Investigation; validation; and writing–review and editing. **Marta Kalousova**: Writing–review and editing and resources. **Jan Kral**: Investigation; validation; and writing–review and editing. **Eva Machackova**: Resources and writing–review and editing. **Barbora Nemcova**: Writing–review and editing and validation. **Marketa Safarikova**: Resources and writing–review and editing. **Drahomira Springer**: Resources and writing–review and editing. **Barbora Stastna**: Investigation; validation; and writing–review and editing. **Spiros Tavandzis**: Writing–review and editing and resources. **Michal Vocka**: Writing–review and editing and resources. **Tomas Zima**: Writing–review and editing and resources. **Jana Soukupova**: Conceptualization; data curation; funding acquisition; methodology; supervision; writing–review and editing; and resources. **Petra Kleiblova**: Conceptualization; funding acquisition; project administration; supervision; writing–review and editing; and resources. **Corinna Ernst**: Formal analysis; methodology; and writing–review and editing. **Zdenek Kleibl**: Conceptualization; funding acquisition; methodology; project administration; supervision; writing–review and editing; and writing–original draft. **Marketa Janatova**: Conceptualization; methodology; writing–original draft; writing–review and editing; data curation; formal analysis; funding acquisition; project administration; supervision; visualization; and resources.

AFFILIATIONS

[1]Institute of Medical Biochemistry and Laboratory Diagnostics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

[2]Institute of Pathological Physiology, First Faculty of Medicine, Charles University, Prague, Czech Republic

[3]Department of Paediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

[4]Department of Genetics and Microbiology, Faculty of Science, Charles University in Prague, Prague, Czech Republic

[5]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic

[6]Centre for Medical Genetics and Reproductive Medicine, GENNET, Prague, Czech Republic

[7]Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic

[8]Department of Medical Genetics, AGEL Research and Training Institute, AGEL Laboratories, Novy Jicin, Czech Republic

[9]Department of Oncology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

[10]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

[11]Centre for Familial Breast and Ovarian Cancer, Center for Integrated Oncology, Medical Faculty, University of Cologne and University Hospital Cologne, Cologne, Germany

ACKNOWLEDGMENTS

This work has been supported by grant projects of the Ministry of Health of the Czech Republic (NU20-09-00355 and DRO-VFN-64165), Charles University projects (COOPERATIO, SVV260631, and UNCE/24/MED/022), and Ministry of Education, Youth and Sports of the Czech Republic Programme EXCELES (LX22NPO5102, funded by the European Union, NextGenerationEU).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available because of privacy or ethical restrictions.

ORCID

Corinna Ernst https://orcid.org/0000-0001-7756-8815
Zdenek Kleibl https://orcid.org/0000-0003-2050-9667
Marketa Janatova https://orcid.org/0000-0003-2816-6769

REFERENCES

1. Ferlay JEM, Lam F, Colombet M, et al. *Global Cancer Observatory: Cancer Today*. International Agency for Research on Cancer; 2022.
2. Mukama T, Kharazmi E, Sundquist K, Sundquist J, Brenner H, Fallah M. Familial risk of breast cancer by dynamic, accumulative, and static definitions of family history. *Cancer*. 2020;126(12):2837-2848. doi:10.1002/cncr.32815
3. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet*. 2008;40(1):17-22. doi:10.1038/ng.2007.53
4. Kleibl Z, Kristensen VN. Women at high risk of breast cancer: molecular characteristics, clinical presentation and management. *Breast*. 2016;28:136-144. doi:10.1016/j.breast.2016.05.006
5. Ghoussaini M, Fletcher O, Michailidou K, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet*. 2012;44(3):312-318. doi:10.1038/ng.1049
6. Gracia-Aznarez FJ, Fernandez V, Pita G, et al. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PLoS One*. 2013;8(2):e55681. doi:10.1371/journal.pone.0055681
7. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447(7148):1087-1093. doi:10.1038/nature05887
8. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013;45(4):353-361. doi:10.1038/ng.2563
9. Sawyer S, Mitchell G, McKinley J, et al. A role for common genomic variants in the assessment of familial breast cancer. *J Clin Oncol*. 2012;30(35):4330-4336. doi:10.1200/JCO.2012.41.7469
10. Mavaddat N, Pharoah PDP, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5):djv036. doi:10.1093/jnci/djv036
11. Reeves GK, Travis RC, Green J, et al. Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. *J Am Med Assoc*. 2010;304(4):426-434. doi:10.1001/jama.2010.1042
12. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet*. 2002;31(1):33-36. doi:10.1038/ng853
13. Pain O, Gillett AC, Austin JC, Folkersen L, Lewis CM. A tool for translating polygenic scores onto the absolute scale using summary statistics. *Eur J Hum Genet*. 2022;30(3):339-348. doi:10.1038/s41431-021-01028-z
14. Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst*. 2010;102(21):1618-1627. doi:10.1093/jnci/djq388
15. Evans DG, Brentnall A, Byers H, et al. The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case-control study. *J Med Genet*. 2017;54(2):111-113. doi:10.1136/jmedgenet-2016-104125
16. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*. 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
17. Brentnall AR, van Veen EM, Harkness EF, et al. A case-control evaluation of 143 single nucleotide polymorphisms for breast cancer risk stratification with classical factors and mammographic density. *Int J Cancer*. 2020;146(8):2122-2129. doi:10.1002/ijc.32541
18. Foretova L, Navratilova M, Svoboda M, et al. Recommendations for preventive care for women with rare genetic cause of breast and ovarian cancer. *Klin Onkol*. 2019;32(suppl 2):6-13. doi:10.14735/amko2019S6
19. Soukupova J, Zemankova P, Lhotova K, et al. Validation of CZECANCA (CZEch CAncer paNel for Clinical Application) for targeted NGS-based analysis of hereditary cancer syndromes. *PLoS One*. 2018;13(4):e0195761. doi:10.1371/journal.pone.0195761
20. Soukupova J, Zemankova P, Kleiblova P, Janatova M, Kleibl Z. CZECANCA: CZEch CAncer paNel for Clinical Application—design and optimization of the targeted sequencing panel for the identification of cancer susceptibility in high-risk individuals from the Czech Republic. *Klin Onkol*. 2016;29(suppl 1):S46-S54. doi:10.14735/amko2016s46
21. Kral J, Jelinkova S, Zemankova P, et al. Germline multigene panel testing of patients with endometrial cancer. *Oncol Lett*. 2023;25(6):216. doi:10.3892/ol.2023.13802
22. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet*. 2015;47(4):373-380. doi:10.1038/ng.3242
23. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551(7678):92-94. doi:10.1038/nature24284
24. Vachon CM, Pankratz VS, Scott CG, et al. The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst*. 2015;107(5):dju397. doi:10.1093/jnci/dju397
25. Shieh Y, Hu D, Ma L, et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res Treat*. 2016;159(3):513-525. doi:10.1007/s10549-016-3953-2
26. Muranen TA, Greco D, Blomqvist C, et al. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. *Genet Med*. 2017;19(5):599-603. doi:10.1038/gim.2016.147
27. Kuchenbaecker KB, McGuffog L, Barrowdale D, et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst*. 2017;109(7):djw302. doi:10.1093/jnci/djw302
28. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet*. 2017;49(5):680-691. doi:10.1038/ng.3826
29. Yang X, Leslie G, Gentry-Maharaj A, et al. Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. *J Med Genet*. 2018;55(8):546-554. doi:10.1136/jmedgenet-2018-105313
30. Lakeman IMM, Hilbers FS, Rodriguez-Girondo M, et al. Addition of a 161-SNP polygenic risk score to family history-based risk prediction: impact on clinical management in non-BRCA1/2 breast cancer families. *J Med Genet*. 2019;56(9):581-589. doi:10.1136/jmedgenet-2019-106072
31. Jia G, Lu Y, Wen W, et al. Evaluating the utility of polygenic risk scores in identifying high-risk individuals for eight common cancers. *JNCI Cancer Spectr*. 2020;4(3):pkaa021. doi:10.1093/jncics/pkaa021
32. Hilbers FS, van Hof PJ, Meijers CM, et al. Clustering of known low and moderate risk alleles rather than a novel recessive high-risk gene in non-BRCA1/2 sib trios affected with breast cancer. *Int J Cancer*. 2020;147(10):2708-2716. doi:10.1002/ijc.33039
33. Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and subtype-specific analyses. *Nat Genet*. 2020;52(6):572-581. doi:10.1038/s41588-020-0609-2
34. Jervis S, Song H, Lee A, et al. Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. *J Med Genet*. 2014;51(2):108-113. doi:10.1136/jmedgenet-2013-102015
35. Lilyquist J, Ruddy KJ, Vachon CM, Couch FJ. Common genetic variation and breast cancer risk—past, present, and future. *Cancer Epidemiol Biomarkers Prev*. 2018;27(4):380-394. doi:10.1158/1055-9965.EPI-17-1144
36. Horackova K, Frankova S, Zemankova P, et al. Low frequency of cancer-predisposition gene mutations in liver transplant candidates with hepatocellular carcinoma. *Cancers*. 2022;15(1):201. doi:10.3390/cancers15010201
37. Borde J, Ernst C, Wappenschmidt B, et al. Performance of breast cancer polygenic risk scores in 760 female CHEK2 germline mutation carriers. *J Natl Cancer Inst*. 2021;113(7):893-899. doi:10.1093/jnci/djaa203
38. Lambert SA, Gil L, Jupp S, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet*. 2021;53(4):420-425. doi:10.1038/s41588-021-00783-5
39. Dusek L, Muzik J, Kubasek M, Koptikova J, Zaloudik J, Vyzula R. *Epidemiology of Malignant Tumours in the Czech Republic*. Masaryk University; 2005. Accessed March 3, 2022. http://www.svod.cz
40. Antoniou AC, Goldgar DE, Andrieu N, et al. A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. *Genet Epidemiol*. 2005;29(1):1-11. doi:10.1002/gepi.20074
41. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM. Time-varying covariates and coefficients in Cox regression models. *Ann Transl Med*. 2018;6(7):121. doi:10.21037/atm.2018.02.12
42. Antoniou AC, Beesley J, McGuffog L, et al. Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction. *Cancer Res*. 2010;70(23):9742-9754. doi:10.1158/0008-5472.CAN-10-1907
43. Duffy SW, Tabar L, Yen AM, et al. Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women. *Cancer*. 2020;126(13):2971-2979. doi:10.1002/cncr.32859
44. Giaquinto AN, Sung H, Miller KD, et al. Breast cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(6):524-541. doi:10.3322/caac.21754
45. Kim G, Bahl M. Assessing risk of breast cancer: a review of risk prediction models. *J Breast Imaging*. 2021;3(2):144-155. doi:10.1093/jbi/wbab001
46. Dite GS, MacInnis RJ, Bickerstaffe A, et al. Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women aged under 50 years: Australian Breast Cancer Family Registry. *Cancer Epidemiol Biomarkers Prev*. 2016;25(2):359-365. doi:10.1158/1055-9965.EPI-15-0838
47. Rudolph A, Song M, Brook MN, et al. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. *Int J Epidemiol*. 2018;47(2):526-536. doi:10.1093/ije/dyx242
48. Tasa T, Puustusmaa M, Tõnisson N, Kolk B, Padrik P. Precision breast cancer screening with a polygenic risk score. *medRxiv*. Preprint posted online August 19, 2020. doi:10.1101/2020.08.17.20176263
49. Jiao Y, Truong T, Eon-Marchais S, et al. Association and performance of polygenic risk scores for breast cancer among French women presenting or not a familial predisposition to the disease. *Eur J Cancer*. 2023;179:76-86. doi:10.1016/j.ejca.2022.11.007
50. Barnes DR, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet Med*. 2020;22(10):1653-1666. doi:10.1038/s41436-020-0862-x
51. Lakeman IMM, van den Broek AJ, Vos JAM, et al. The predictive ability of the 313 variant-based polygenic risk score for contralateral breast cancer risk prediction in women of European ancestry with a heterozygous BRCA1 or BRCA2 pathogenic variant. *Genet Med*. 2021;23(9):1726-1737. doi:10.1038/s41436-021-01198-7
52. Evans DGR, van Veen EM, Harkness EF, et al. Breast cancer risk stratification in women of screening age: incremental effects of adding mammographic density, polygenic risk, and a gene panel. *Genet Med*. 2022;24(7):1485-1494. doi:10.1016/j.gim.2022.03.009
53. Yiangou K, Mavaddat N, Dennis J, et al. Differences in polygenic score distributions in European ancestry populations: implications for breast cancer risk prediction. *medRxiv*. Preprint posted online February 13, 2024. doi:10.1101/2024.02.12.24302043
54. Lakeman IMM, Rodriguez-Girondo MDM, Lee A, et al. Clinical applicability of the polygenic risk score for breast cancer risk prediction in familial cases. *J Med Genet*. 2023;60(4):327-336. doi:10.1136/jmg-2022-108502
55. Mavaddat N, Ficorella L, Carver T, et al. Incorporating alternative polygenic risk scores into the BOADICEA breast cancer risk prediction model. *Cancer Epidemiol Biomarkers Prev*. 2023;32(3):422-427. doi:10.1158/1055-9965.EPI-22-0756
56. Hughes E, Tshiaba P, Gallagher S, et al. Development and validation of a clinical polygenic risk score to predict breast cancer risk. *JCO Precis Oncol*. 2020;4:585-592. doi:10.1200/PO.19.00360
57. Bolze A, Cirulli ET, Hajek C, Schnell Blitstein JM, Grzymski JJ. The potential of genetics in identifying women at lower risk of breast cancer. *JAMA Oncol*. 2024;10(2):236-239. doi:10.1001/jamaoncol.2023.5468
58. Mars N, Kerminen S, Tamlander M, et al. Comprehensive inherited risk estimation for risk-based breast cancer screening in women. *J Clin Oncol*. Published online February 29, 2024. doi:10.1200/jco.23.00295

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hovhannisyan M, Zemankova P, Nehasil P, et al. Population-specific validation and comparison of the performance of 77- and 313-variant polygenic risk scores for breast cancer risk prediction. *Cancer*. 2024;130(17):2978-2987. doi:10.1002/cncr.35337

10970142, 2024, 17, Downloaded from https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.35337 by Charles University, Wiley Online Library on [22/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License