Combination of phenotype and polygenic risk score in breast cancer risk evaluation in the Spanish population

Juan Trivino
Sistemas Genomicos
A Ceba
Sistemas Genomicos
E Rubio-Solsona
Sistemas Genomicos
D Serra
Sistemas Genomicos
I Sanchez-Guiu
Sistemas Genomicos
G Ribas
Sistemas Genomicos
R Rosa
Sistemas Genomicos
M Cabo
Sistemas Genomicos
L Bernad
Sistemas Genomicos
G Pita
Spanish National Genotyping Center
A Gonzalez-Neira
Spanish Natinal Genotyping Center
G Legarda
Sistemas Genomicos
JL Diaz
Sistemas Genomicos
A García-Vigara
Obstretrics and Gynecology Service
A Martínez-Aspas
Obstretics and Gynecology Service. INCLIVA
M Escrig
Departament of Hematology and Medical Oncoloy. INCLIVA
B Bermejo
Departament of Hematology and Medical Oncology.INCLIVA
P Eroles
Departament of Hematology and Medical Oncology.INCLIVA
J Ibáñez

General Directorate Public Health, Valencian Community.

D Salas

General Directorate Public Health, Valencian Community

A Julve

Radiology Service.Hospital Clinico Universitario. Valencia

A Cano

Obstetrics and Gynecology Service.Hospital Clinico Universitario.Valencia

A Lluch

Ciberonc. Incliva

R Miñambres

Sistemas Genomicos

Javier Benitez (jbenitez@cnio.es)

Spanish National Genotyping Center

Research article

Keywords: Polygenic risk score; predictive test; identification of high risk women; risk algorithms

Posted Date: August 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-47631/v1

Version of Record: A version of this preprint was published on November 10th, 2020. See the published version at https://doi.org/10.1186/s12885-020-07584-9.

Abstract

Background: In the last years, the identification of different genetic and phenotypic biomarkers for prevention, early diagnostics and patient stratification have been a main objective in cancer research. Different proposals for multivariable models using biomarkers have been presented for the evaluation of individual risk of women to develop breast cancer.

Methods: In this study, we describe and evaluate a multivariable model using 92 Single-nucleotide polymorphisms (SNPs) and five different phenotypic variables in a Spanish population of 642 healthy women and 455 breast cancer patients.

Results: We were able to stratify both groups with our model. The 9th decile included 1% of controls vs 9% of cases, with an Odds ratio (OR) of 12.9 and a p-value of 3.43E-07**.** The first decile presented an inverse proportion: 1% of cases and 9% of controls, with an OR of 0.097 and a p-value of 1.86E-08.

Conclusions: These results indicate the ability of the multivariable model to stratify women according to their risk to develop breast cancer over a maximum period of 5 years. The analysis present a proof of concept in a poorly studied population and it opens the possibility of using this type of method for routine screening in the Spanish population.

Background

Prevention and early diagnosis of breast cancer is one of the main objectives in cancer research. There are different models to estimate cancer risk that are based on non-genetic or genetic factors, that is, high or moderate susceptibility genes [1, 2]. However, the extensive use of genome-wide association studies (GWAS) has promoted the use of low susceptibility alleles, which in combination with non-genetic factors can modify the risk of developing breast cancer [3]. We recently described a 76 SNP polygenic risk score (PRS) for breast cancer that permitted to stratify the general population. Women at low and high risk of developing breast cancer presented 0.5 and 2.5-fold increased risks, respectively, relative to women in the middle quintile [4]. Therefore, the combination of this PRS with environmental and personal factors would increase our ability to identify women at risk that would be candidates for a personalized follow-up. Previous studies have shown that breast density or familial antecedents and PRS models composed of 77 [5], 83 [6] or, more recently, 313 SNPs [7] discriminate women at risk. The combination of phenotype and PRS improves discrimination considerably, indicating that PRS plays a role in risk stratification. In addition, the combination of non-genetic and genetic factors could guide screening strategies when an individual exceeds the risk threshold.

Although there are some previous studies in Caucasian populations, this is the first study in the Spanish population combining a PRS of 92 SNPs with other risk factors such as mammographic density (MD), reproductive factors, and family history in order to validate previous results and analyze the usefulness of this type of studies.

Methods

Study design: description of cohorts

The present study was submitted to and approved by the Clinical Research Ethics Committee (CEIC) of the Hospital Clínico Universitario de Valencia (Spain), September 29th, 2016 (2016/169) and July 13th, 2018 (2018/139), and conducted in compliance with the Helsinki Declaration.

The case control study with full genotyping and phenotypic data include a cohort from two centers: Clinical University Hospital of Valencia and Valencian Community Screening Programme (General Directorate Public Health), both from the Autonomous Community of Valencia (Mediterranean Coast), with a total of 867 healthy women and 640 patients. The cases were described as women who development breast cancer over a maximum period of five years from data collections. On other hands, controls were those women who did not development breast cancer during the same period. Finally, women that presented incomplete phenotypic data or genotyping failure from cohort were excluded reaching a final case control group of 1,097 participants consisting of 642 healthy women and 455 cases with an age range of 30 and 75 years old.

Data collection

Clinical information of all subjects was collected at recruitment: family history of breast cancer, birth date, age (age at diagnosis for cases), age at menarche, age at menopause, age at first pregnancy, and mammographic density (MD).

Breast density was assessed from the craniocaudal and mediolateral oblique mammographic projections by a single, experienced radiologist. In healthy women, only the densest breast was evaluated, while in cases the tumor-free breast was considered. The radiologist classified MD according to Boyd's semiquantitative scale into 5 categories of density (0–10%; 10–25%; 25–50%; 50–75%: >75%) [8].

SNP selection and genotyping

For our previous PRS risk analysis, 76 SNPs were initially selected from the European Collaborative Oncological Gene Environment Study (COGS)[9]. These SNPs were significant or with a trend to significance in our previous validation with Spanish samples. The correlation of the genetic variants analyzed with the prediction of breast cancer risk in women of the Spanish population has been already described [4]. Briefly, in this study is described the performance of our PRS using the 76 selected SNPs in breast cancer risk prediction of women in a case and control Spanish cohort. The initial selection was extended to 123 SNPs by including additional SNPs obtained from the OncoArray Project (10). From these, 28 SNPs with an OR close to 1 ($0.95 \leq OR \leq 1.05$) and another 3 SNPs with platform genotyping failure were removed. Finally, a total of 92 SNPs [11–16] was used for further analysis (Online Resource 1).

The genotyping method has been previously described [4]. Briefly, 10 ml of peripheral blood was collected in an EDTA tube. One µg of Deoxyribonucleic acid (DNA) was used for the genotypic analysis (minimum concentration of 25 ng/µL). Genotyping was performed with the Open Array® Real-Time PCR platform (Life Technologies) using the Acufill® system and Taqman® probes. Data obtained were analyzed using Genotyper software. Samples with a call rate < 0.95 were discarded. SNPs with a genotyping rate < 0.95 and SNPs with errors in control duplicates were also ruled out.

Statistical analysis

In an initial exploratory univariable process, the cases/controls ratio of each risk factor was compared. In this step, the Wilcoxon-test was used with a two-sided p-value threshold of 0.05.

The PRS was based on a combined effect of 92 SNPs statistically associated with breast cancer. This strategy considers an independent effect of each SNP, ignoring departures from a multiplicative model [17]. The PRS was derived for each study subject using the formula:(see Formula 1 in the Supplementary Files)

Where xk is the number of risk alleles (0, 1 or2) based on the ploidy of each SNP. The βk weights are the ORs of the risk alleles associated with breast cancer described in Online Resource 1. This strategy has been used in other studies [5, 6]. The resulting values are normalized using the median PRS value of the control samples of the cohort.

Regarding phenotypic analysis, the phenotypic categories were transformed into quantitative variables using the ORs described in the *Pollan et al.* study [8], except for family history; the ORs of this latter category were based on the *Pharoah et al.* study [18]. In addition, the reference age of women (age at diagnosis for cases and age at interview for controls) was grouped into five-year periods, similar to other publications [19], then, the groups were transformed into a quantitative variable. The final number of cases and control of our cohort were 455 and 642 respectively.

For the univariable analysis, logistic regression was applied to each risk factor adjusted for age and centre. The coefficients of the model were standardized using the reghelper [20] library of R. Additionally; the PRS was adjusted by the first five principal components. The interaction effect between variables was also evaluated using the likelihood ratio test (LRT). All analyses were two-sided using a p-value threshold of 0.05.

To check the assumption of independence of the PRS and other phenotypic risk factors, pairwise Spearman correlations using unaffected controls were evaluated.

Regarding the multivariable study, the analysis was based on logistic regression incorporating the statistically significant variables obtained in the previous steps, including the interaction terms. Family history and age at menarche, although not significant, were also included in the analyses since they are well-known risk factors. The significance of the final model was evaluated using the Wald Test [21]. To assess the accuracy of the final multivariable model, a global Hosmer-Lemeshow goodness of fit test was performed using deciles [22].

To evaluate improvement in risk prediction for the different models and risk factors, the area under the curve (AUC) was evaluated [23] as a measure of discrimination between cases and control women. This calculation was performed using the pROC [24] library of R. To avoid a possible overfit adjustment of the model, the 95% Confidence Interval (CI) of the AUC was assessed using a cross validation strategy [25]. This step was based on the calculation of AUC in 1,000 permutations using a random selection of 90% of women as training set and the remaining 10% as test set.

Finally, women were stratified into deciles based on their final individual risk factor obtained from the multivariable model. The ORs of extreme deciles were evaluated using logistic regression with the 40–60% range as reference.

Based on the characteristics of our cohort, the final individual risk factor proposed in this study describes the relative risk of women of the Spanish population to suffer breast cancer in a maximum period of five years.

Results

Association of phenotypic risks factors with breast cancer

The age of women is one of the most important risk factors of breast cancer [26]. To ensure that no bias or confounding effect associated with this risk factor affected the analysis, the case and control distributions were compared. No significant differences were found using the Wilcoxon test (p-value of 0.27). The median age of our cohort was 51 years old with a range of 39 and 61 years in the extreme deciles (Table 1).

Table 1
Phenotypic and genotypic baseline characteristics of cases and controls in our Spanish cohort.

Risk Factor	Category	Description	Number Controls	% Controls	Number Cases	% Cases	OR	OR CI 95%	p-value
Age	0	30–35 years	28	4.36	16	3.52	1.05	0.79–1.13	0.27
	1	35–40 years	58	9.03	28	6.15			
	2	40–45 years	84	13.08	63	13.85			
	3	45–50 years	138	21.50	115	25.27			
	4	50–55 years	158	24.61	86	18.90			
	5	55–60 years	113	17.60	94	20.66			
	6	60–65 years	47	7.32	37	8.13			
	7	> 65 years	16	2.49	16	3.52			
Breast Density	0	From 0 to 10%	99	15.42	51	11.21	1.46	1.21–1.71	1.64E-07
	1	From 11 to 25%	116	18.07	53	11.65			
	2	From 26 to 50%	185	28.82	116	25.49			
	3	From 51 to 75%	181	28.19	133	29.23			
	4	Greater than 75%	61	9.50	102	22.42			
Age at first delivery	0	Less than 20 years	33	5.14	23	5.05	1.15	1.02–1.31	0.03
	1	From 20 to 24 years	165	25.70	104	22.86			
	2	From 25 to 29 years	203	31.62	107	23.52			
	3	From 30 to 34 years	106	16.51	101	22.20			
	4	Greater than 34 years	56	8.72	50	10.99			
	5	Nulliparous	79	12.31	70	15.38			
Age at menopause	0	Less than 46 years	97	15.11	47	10.33	1.96	1.72–2.24	2.20E-16
	1	From 46 to 50 years	147	22.90	102	22.42			
	2	Greater than 50 years	110	17.13	71	15.60			
	3	Premenopause	87	13.55	212	46.59			

Risk Factor	Category	Description	Number Controls	% Controls	Number Cases	% Cases	OR	OR CI 95%	p-value
	4	Menstruating	201	31.31	23	4.97			
Age at menarche	0	Equal to or greater than 15 years	34	5.30	34	7.47	0.89	0.78–1.04	0.061
	1	14 years	115	17.91	85	18.68			
	2	13 years	178	27.73	100	21.98			
	3	12 years	140	21.81	110	24.18			
	4	Less than 12 years	175	27.26	123	27.03			
	5	Null	0	0.00	3	0.66			
Familial antecedents	0	No affected relative	468	72.90	308	67.69	1.05	0.93–1.19	0.34
	1	A first-degree relative diagnosed with breast cancer at age 50 years or older	52	8.10	43	9.45			
	2	A first-degree relative diagnosed with breast cancer younger than 50 years	25	3.89	18	3.96			
	3	1 affected second-degree relative	90	14.02	79	17.36			
	4	2 affected first-degree relatives	4	0.62	5	1.10			
	5	2 affected second-degree relatives	1	0.16	2	0.44			
	6	3 or more affected relatives	2	0.31	0	0.00			

The global phenotypic risk factors after comparison between cases and controls of the cohort are detailed in Table 1. Age at menarche and familial antecedents were not statistically significantly different between cases and controls, with p-values of 0.061 and 0.34, respectively.

Mammographic density presented a clear statistically significant relationship with breast cancer, with an OR of 1.46 (95% CI: 1.21–1.71) and a p-value of 1.64E-7. The main differences between controls and cases are concentrated in the extremes, with respective proportions of 15% versus 11% in the first category (MD 0–10%) and 10% versus 22% in the last category (MD > 75%).

In this study, a higher age at first delivery was associated to an increase risk of development of breast cancer, while age at menarche had no statistically significant effect; the p-values were 0.03 and 0.061 respectively. The age at first delivery was associated with an OR of 1.15 (95% CI: 1.02–1.31) and the main difference was seen at advanced maternal ages (greater than 34 years) with a proportion of 11% versus 8% in cases and controls, respectively. Regarding the age at menarche, the OR was 0.89 (95% CI: 0.78–1.04). Another reproductive factor considered in this study was the menopause status, which was associated with an OR of 1.96 (95% CI: 1.72–2.24) and a p-value < 2E-16. The highest difference between cases and controls was observed in the premenopausal category, with values of 46% and 13% respectively. Regarding the family history, cases showed a slight trend to present more breast cancer antecedents in first- and second-degree family members than controls; however, the logistic regression based on this quantitative variable was not statistically significant with a p-value of 0.34.The interaction terms identified in our analysis as statistically significant and included in the multivariable model were the age of women with mammary density and age with menopause status, with p-values of 0.004 and 2E-16, respectively. The relation between both phenotype variables and the age of women has been studied for a long time in the breast cancer context [27–29].

The discriminative power of each phenotypic risk factor was compared using ROC curve analysis generated by 10-fold cross-validation (Table 2). The results were concordant with the univariable logistic regression, where age at menarche and family history did not present significant trends and the most discriminant phenotypic variables were menopause status with an AUC of 0.64 (95% CI: 0.58–0.70) and mammographic density with an AUC of 0.60 (95% CI: 0.56–0.66) (Table 2).

Table 2
Age-adjusted AUC for univariable and multivariable models.

Model	Median AUC	95% CI AUC	p-value
Breast Density	0.60	0.54–0.66	2.17E-03
Age at first delivery	0.54	0.48–0.60	1.49E-01
Age at Menopause	0.64	0.58–0.70	5.40E-09
Familial Antecedents	0.52	0.47–0.58	6.45E-01
Age at Menarche	0.53	0.48–0.59	2.80E-01
PRS92	0.62	0.56–0.66	3.64E-03
Multivariable model no interactions	0.74	0.71–0.77	2.20E-16
Multivariable model with interactions	0.8	0.77–0.83	2.20E-16

Association of PRS92 with breast cancer

The PRS based on 92 SNPs presented an OR per 1 standard deviation (SD) of 1.41 with a 95% CI of 1.24–1.61 and a p-value of 6.30×10^{-8}.For women in the lowest quintile (5%), the PRS distribution presented an OR of 0.38 (95% CI: 0.22–0.63; p-value = 0.0026) compared with women in the middle quintile (40–60%). On the other hand, the highest quintile (95%) PRS distribution exhibited an OR of 1.87 (95% CI: 1.16 3.08; p = 0.036) (Fig. 1). The x-axis shows the different deciles and the y-axis the OR using the 40–60% range as reference. The discriminative accuracy of the

PRS92 was calculated using the area under the curve (AUC). The PRS92 adjusted by age and the first five principal components presented a discriminative power of 0.62 with a 95% CI of 0.56–0.66 (Table 2). This predictive performance range was, along with breast mammographic density (0.60) and menopause (0.64), one of the most discriminant variables.

Multivariable model for breast cancer stratification

All statistically significant univariable risk factors and interaction terms were included in the final multivariable model. Age at menarche and family history were also incorporated based on the scientific literature. No other variable presented significant correlations using the Spearman method (data not shown).

The discriminative accuracy of the multivariable model with and without interaction terms was evaluated (Fig. 2). The median AUC obtained using the interaction model was 0.80 (95% CI: 0.77–0.83) and was higher than in the model without interactions, 0.74 (95% CI: 0.71–0.77) (Table 2). This difference was statistically significant with a p-value of 5.375E-09. These values are slightly higher than those observed in other published strategies [30, 31].

We investigated how the individual risk for cases and controls differed when the final multivariable model was used. (Fig. 3) and (Table 3) show the ORs and the percentages of cases and controls classified by different deciles using the final risk predicted by the multivariable model with interactions. In the first decile, the OR was 0.097 (CI: 95% 0.046–0.184) with a p-value of 1.86E-08. In this range are contained 9% of controls versus less than 1% of cases. This trend is similar in the next decile, with 8.75% and 1.28% of controls and cases, respectively (OR: 0.29; p = 8.12E-07). At the other extreme, the last decile presented an OR of 12.9 (CI: 95% 5.098–23.332; p = 3.43E-07). In this decile, the proportion of cases and control was inversed, with 9% of cases and 1% of controls. These results indicate the ability of the multivariable model to stratify women according to the risk to suffer breast cancer over a maximum period of 5 years.

Table 3
ORs, 95% CI and distribution of cases and controls in deciles

Deciles	OR	OR 5%	OR95%	p-value	% Controls	%Cases
< 10%	0.097	0.046	0.184	1.86E-08	9.39	0.64
10–20%	0.209	0.121	0.345	8.12E-07	8.75	1.28
20–40%	0.402	0.282	0.570	1.99E-05	27.35	8.20
60–80%	1.803	1.313	2.481	2.30E-03	8.84	11.12
80–90%	3.071	2.057	4.634	5.31E-06	3.19	6.84
> 90%	12.900	5.098	23.332	3.43E-07	1.00	9.02
Results using the multivariable model with interactions. The 40–60% range was selected as reference.						

Discussion

In the last years, different proposals for multivariable models have been presented that were centered on the stratification of women who might suffer breast cancer based on their individual risk. Different biomarkers have been

analyzed as possible predictors, including phenotypic and non-phenotypic markers, and environmental and genetic factors.

Regarding the genetic variables, one approach uses the polygenic risk score (PRS). This strategy is based on the amount of a variable number of statistically significant low penetrance variants obtained from large GWAS analyses [5, 32].

Our study was based on one relatively small cohort of women. We adjusted for the origin center of the women in our univariable and multi-variable models.

We used a specific PRS based on 92 SNPs and we obtained an OR of 1.41 (1.24–1.61) that was consistent with other published studies based on Caucasian populations using different numbers of SNPs (from 18 up to 313 SNPs) [5, 32–34].

In addition, the AUC-ROC was 0.62 with a 95% CI of 0.56–0.66, consistent with the literature, where the range assigned to European populations was from 0.58 to 0.65 and from 0.53 to 0.64 for non-European populations [35].

Regarding univariable phenotypic risk factor analysis, the most statistically significant result is based on the discriminant variables identified, the menopause status and mammographic density, which is consistent with other studies [29, 36–38]. Other reproductive factors such as later age at first birth and later age at menarche have been identified as risk factors for breast cancer [39]. In our results, a significant p-value of 0.03 and an OR of 1.15 were identified for the first risk factor. Age at menarche was not found not to be statistically significant (p-value = 0.061).

The ORs of risk factors obtained in our cohort presents some differences respect to others previously reported results. The most evident are the non-statistically significance of family history and the age of menarche. However, the direction of these well-established effects and our results are concordant. On other hands, the magnitude of OR in mammographic density were lower respect the previous literature. These differences may be due to the low number of women of our cohort and new women may be incorporated in future works, however, the concordance in the effect, direction and magnitude of the different ORs found in our cohort corroborates the usefulness of our study as a first proof of concept in Spanish population

Additionally, the joint association of our PRS92 with transformed continuous phenotypic variables, such as MD, reproductive factors and family history was examined in our Spanish population. We found that the genotypic and phenotypic variables did not present any significant correlation between them; possibly, a multiplicative model can describe this better and help to improve breast cancer risk estimation.

The precision of the multivariable model increased when adding two statistically significant interaction terms associated with the age of women, menopause and mammographic density. These interactions have been previously observed and we found an increase of AUC-ROC from 0.74 (95% CI: 0.71–0.77) to 0.80 (95% CI: 0.77–0.83) (Table 2). This increase was statistically significant and offered a final value slightly higher than other similar multivariable studies [40].

We were able to stratify the control group with our model (Fig. 3). Both extremes showed important differences. The last decile included 1% of controls vs 9% of cases, with an OR of 12.9 and a p-value 3.43E-07. By contrast, the first decile presented an inverse proportion (1% of cases and 9% of controls); in this case, the OR was 0.097 with a p-value of 1.86E-08. These results indicate the ability of the multivariable model to stratify women according to their risk to suffer breast cancer over a maximum period of 5 years.

In summary, the results indicate that the multivariable logistic model using a combination of genetic, phenotype and interaction variables is useful for stratification of women according to their of individual risk to suffer breast cancer in a 5-year period in the Spanish population; this capacity is similar to that observed in other studies in European and non-European populations. The major limitation of our analysis is the small size of the cohort; however, the importance of our analysis is that it is a proof of concept in a population never studied. Larger series are necessary in order to confirm our data and start to use this type of method for screening in the Spanish population.

Conclusions

These results indicate the ability of the multivariable model to stratify women according to their risk to develop breast cancer over a maximum period of 5 years. The analysis present a proof of concept in a poorly studied population and it opens the possibility of using this type of method for routine screening in the Spanish population.

Abbreviations

CNIO: Spanish National Cancer Center

SNP: Single-Nucleotide Polymorphism

OR: Odd Ratio

GWAS: Genome-Wide Association Studies

PRS: Polygenic Risk Score

MD: Mammographic Density

COGS: European Collaborative Oncological Gene Environment Study

DNA: Deoxyribonucleic acid

LRT: Likelihood Ratio Test

AUC: Area Under the Curve

CI - Confidence Interval

SD - Standard Deviation

Declarations

Ethics approval and consent to participate

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The present study was submitted to and approved by the Clinical Research Ethics Committee (CEIC) of the Hospital Clínico Universitario de Valencia (Spain), September 29th, 2016 (2016/169) and July 13th, 2018 (2018/139), and conducted in compliance with the Helsinki Declaration.

Informed consent: Written informed consent was obtained from all individual participants included in the study at the time of recruitment. Patient information was anonymized and de-identified.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by own funds of Sistemas Genómicos S.L. Additionally, the study was supported by two funders that provided funds for the execution of the project: IVACE (Instituto Valenciano de competitividad empresarial) and AVI (Agencia Valenciana de Innovación). IVACE grant numbers: IMIDTA/2016/75 and IMIDTA/2018/7; AVI programme grant number; INNTAL21/197002.

Authors' contributions

JCT, AC, ER, RM, RB, AGN, AGV, AMA, BB, PE, JI, DS conceived experiments. DS, ISG, GR, GP, SD, AJ, AL carried out experiments. JCT, AC, MC, GL, RR, JLD analysed data. JCT, AC, RM, JB, LB, AL, AC conceived the study, participated in its design and coordination and wrote the manuscript.

Acknowledgements

We acknowledge to Spanish National Genotyping Center's lab (CEGEN) for its technical support and Pablo Fernandez from the Epidemiology Center (Madrid) for his statistical support.

References

[1] Collaborative Group on Hormonal Factors in Breast Cancer (2001). Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet (London, England), 358(9291), 1389–1399. https://doi.org/10.1016/S0140-6736(01)06524-2

[2] Antoniou, A. C., & Easton, D. F. (2006). Models of genetic susceptibility to breast cancer. Oncogene, 25(43), 5898–5905. https://doi.org/10.1038/sj.onc.1209879

[3] Rudolph, A., Song, M., Brook, M. N., Milne, R. L., Mavaddat, N., Michailidou, K., Bolla, M. K., Wang, Q., Dennis, J., Wilcox, A. N., Hopper, J. L., Southey, M. C., Keeman, R., Fasching, P. A., Beckmann, M. W., Gago-Dominguez, M., Castelao, J. E., Guénel, P., Truong, T., Bojesen, S. E et al. (2018). Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. International journal of epidemiology, 47(2), 526–536. https://doi.org/10.1093/ije/dyx242

[4] Miñambres R., Pita G., Sánchez I., Serra D., Rosa R., Rubio-Solsona E., Palacios S., Llaneza A., Marron P., Hoyas S., Lluch A., Cano A., González A., Triviño JC., Benítez J (2019) Prediction of cancer risk based on study of genetic variants in healthy women in the Spanish population. Revista de Senología y Patología Mamaria, Vol 32 (3): 94-99.https://doi.org/10.1016/j.senol.2019.07.001

[5] Mavaddat, N., Pharoah, P. D., Michailidou, K., Tyrer, J., Brook, M. N., Bolla, M. K., Wang, Q., Dennis, J., Dunning, A. M., Shah, M., Luben, R., Brown, J., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Czene, K., Darabi, H., Eriksson, M., Peto J et al (2015). Prediction of breast cancer risk based on profiling with common genetic variants. Journal of the National Cancer Institute, 107(5), djv036. https://doi.org/10.1093/jnci/djv036

[6] Vachon, C. M., Scott, C. G., Tamimi, R. M., Thompson, D. J., Fasching, P. A., Stone, J., Southey, M. C., Winham, S., Lindström, S., Lilyquist, J., Giles, G. G., Milne, R. L., MacInnis, R. J., Baglietto, L., Li, J., Czene, K., Bolla, M. K., Wang, Q., Dennis, J., Haeberle, L et al (2019). Joint association of mammographic density adjusted for age and body mass index and polygenic risk score with breast cancer risk. Breast cancer research: BCR, 21(1), 68. https://doi.org/10.1186/s13058-019-1138-8

[7] Lee, A., Mavaddat, N., Wilcox, A. N., Cunningham, A. P., Carver, T., Hartley, S., Babb de Villiers, C., Izquierdo, A., Simard, J., Schmidt, M. K., Walter, F. M., Chatterjee, N., Garcia-Closas, M., Tischkowitz, M., Pharoah, P., Easton, D. F., & Antoniou, A. C. (2019). BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genetics in medicine : official journal of the American College of Medical Genetics, 21(8), 1708–1718. https://doi.org/10.1038/s41436-018-0406-9

[8] Pollán, M., Ascunce, N., Ederra, M., Murillo, A., Erdozáin, N., Alés-Martínez, J., & Pastor-Barriuso, R. (2013). Mammographic density and risk of breast cancer according to tumor characteristics and mode of detection: a Spanish population-based case-control study. Breast cancer research : BCR, 15(1), R9. https://doi.org/10.1186/bcr3380

[9] Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., Schmidt, M. K., Chang-Claude, J., Bojesen, S. E., Bolla, M. K., Wang, Q., Dicks, E., Lee, A., Turnbull, C., Rahman, N., Breast and Ovarian Cancer Susceptibility Collaboration, Fletcher, O., Peto, J., Gibson, L., Dos Santos Silva, I et al (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nature genetics, 45(4), 353–361e3612. https://doi.org/10.1038/ng.2563

[10] Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M. J., Maranian, M. J., Bolla, M. K., Wang, Q., Shah, M., Perkins, B. J., Czene, K., Eriksson, M., Darabi, H., Brand, J. S., Bojesen, S. E., Nordestgaard, B. G., Flyger, H., Nielsen, S. F., Rahman, N et al (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nature genetics, 47(4), 373–380. https://doi.org/10.1038/ng.3242

[11] Ghoussaini, M., Pharoah, P., & Easton, D. F. (2013). Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning?. The American journal of pathology, 183(4), 1038–1051. https://doi.org/10.1016/j.ajpath.2013.07.003

[12] Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M. K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., French, J. D et al (2017). Association analysis identifies 65 new breast cancer risk loci. Nature, 551(7678), 92–94. https://doi.org/10.1038/nature24284

[13] Osorio, A., Milne, R. L., Kuchenbaecker, K., Vaclová, T., Pita, G., Alonso, R., Peterlongo, P., Blanco, I., de la Hoya, M., Duran, M., Díez, O., Ramón Y Cajal, T., Konstantopoulou, I., Martínez-Bouzas, C., Andrés Conejero, R., Soucy, P., McGuffog, L., Barrowdale, D., Lee, A., SWE-BRCA et al (2014). DNA glycosylases involved in base excision repair may be associated with cancer risk in BRCA1 and BRCA2 mutation carriers. PLoS genetics, 10(4), e1004256. https://doi.org/10.1371/journal.pgen.1004256

[14] Gaudet, M. M., Kuchenbaecker, K. B., Vijai, J., Klein, R. J., Kirchhoff, T., McGuffog, L., Barrowdale, D., Dunning, A. M., Lee, A., Dennis, J., Healey, S., Dicks, E., Soucy, P., Sinilnikova, O. M., Pankratz, V. S., Wang, X., Eldridge, R. C., Tessier, D. C., Vincent, D., Bacot, F et al (2013). Identification of a BRCA2-specific modifier locus at 6p24 related to breast cancer risk. PLoS genetics, 9(3), e1003173. https://doi.org/10.1371/journal.pgen.1003173

[15] Couch, F. J., Wang, X., McGuffog, L., Lee, A., Olswold, C., Kuchenbaecker, K. B., Soucy, P., Fredericksen, Z., Barrowdale, D., Dennis, J., Gaudet, M. M., Dicks, E., Kosel, M., Healey, S., Sinilnikova, O. M., Lee, A., Bacot, F., Vincent, D., Hogervorst, F. B., Peock, S et al (2013). Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS genetics, 9(3), e1003212. https://doi.org/10.1371/journal.pgen.1003212

[16] Bojesen, S. E., Pooley, K. A., Johnatty, S. E., Beesley, J., Michailidou, K., Tyrer, J. P., Edwards, S. L., Pickett, H. A., Shen, H. C., Smart, C. E., Hillman, K. M., Mai, P. L., Lawrenson, K., Stutz, M. D., Lu, Y., Karevan, R., Woods, N., Johnston, R. L., French, J. D., Chen, X et al (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nature genetics, 45(4), 371–384e3842. https://doi.org/10.1038/ng.2566

[17] Pharoah, P. D., Antoniou, A. C., Easton, D. F., & Ponder, B. A. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. The New England journal of medicine, 358(26), 2796–2803. https://doi.org/10.1056/NEJMsa0708739

[18] Pharoah PD., Day NE., Duffy S., Easton DF., Ponder BA. (1997) Family history and the risk of breast cancer: a systematic review and metaanalysis. Int J Cancer;71(5):800-9.

[19] Maas, P., Barrdahl, M., Joshi, A. D., Auer, P. L., Gaudet, M. M., Milne, R. L., Schumacher, F. R., Anderson, W. F., Check, D., Chattopadhyay, S., Baglietto, L., Berg, C. D., Chanock, S. J., Cox, D. G., Figueroa, J. D., Gail, M. H., Graubard, B. I., Haiman, C. A., Hankinson, S. E., Hoover, R. N et al (2016). Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA oncology, 2(10), 1295–1302. https://doi.org/10.1001/jamaoncol.2016.1025

[20] Jeffrey Hughes (2020). reghelper: Helper Functions for Regression Analysis. R package version 0.3.5.

[21] Draper, N.R., Smith, H., (1998). Applied Regression Analysis. New York, John Wiley & Sons

[22] Hosmer DW, Lemeshow S, Sturdivant RX., (2013) Applied logistic regression. 3rd ed. New York: Wiley

[23] Rosner, B., & Glynn, R. J. (2009). Power and sample size estimation for the Wilcoxon rank sum test with application to comparisons of C statistics from alternative prediction models. Biometrics, 65(1), 188–197. https://doi.org/10.1111/j.1541-0420.2008.01062.x

[24] Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC bioinformatics, 12, 77. https://doi.org/10.1186/1471-

2105-12-77

[25] Tworoger, S. S., Zhang, X., Eliassen, A. H., Qian, J., Colditz, G. A., Willett, W. C., Rosner, B. A., Kraft, P., & Hankinson, S. E. (2014). Inclusion of endogenous hormone levels in risk prediction models of postmenopausal breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 32(28), 3111–3117. https://doi.org/10.1200/JCO.2014.56.1068

[26] Siegel, R. L., Miller, K. D., & Jemal, A. (2017). Cancer Statistics, 2017. CA: a cancer journal for clinicians, 67(1), 7–30. https://doi.org/10.3322/caac.21387

[27] Bertrand, K. A., Tamimi, R. M., Scott, C. G., Jensen, M. R., Pankratz, V., Visscher, D., Norman, A., Couch, F., Shepherd, J., Fan, B., Chen, Y. Y., Ma, L., Beck, A. H., Cummings, S. R., Kerlikowske, K., & Vachon, C. M. (2013). Mammographic density and risk of breast cancer by age and tumor characteristics. Breast cancer research : BCR, 15(6), R104. https://doi.org/10.1186/bcr3570

[28] Tice, J. A., Miglioretti, D. L., Li, C. S., Vachon, C. M., Gard, C. C., & Kerlikowske, K. (2015). Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 33(28), 3137–3143. https://doi.org/10.1200/JCO.2015.60.8869

[29] Hsieh, C. C., Trichopoulos, D., Katsouyanni, K., & Yuasa, S. (1990). Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. International journal of cancer, 46(5), 796–800. https://doi.org/10.1002/ijc.2910460508

[30] Shieh, Y., Hu, D., Ma, L., Huntsman, S., Gard, C. C., Leung, J., Tice, J. A., Ziv, E., Kerlikowske, K., & Cummings, S. R. (2017). Joint relative risks for estrogen receptor-positive breast cancer from a clinical model, polygenic risk score, and sex hormones. Breast cancer research and treatment, 166(2), 603–612. https://doi.org/10.1007/s10549-017-4430-2

[31] Dite, G. S., MacInnis, R. J., Bickerstaffe, A., Dowty, J. G., Allman, R., Apicella, C., Milne, R. L., Tsimiklis, H., Phillips, K. A., Giles, G. G., Terry, M. B., Southey, M. C., & Hopper, J. L. (2016). 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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 25(2), 359–365. https://doi.org/10.1158/1055-9965.EPI-15-0838

[32] Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J. P., Chen, T. H., Wang, Q., Bolla, M. K., Yang, X., Adank, M. A., Ahearn, T., Aittomäki, K., Allen, J., Andrulis, I. L., Anton-Culver, H., Antonenkova, N. N., Arndt, V., Aronson, K. J et al (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. American journal of human genetics, 104(1), 21–34. https://doi.org/10.1016/j.ajhg.2018.11.002

[33] Van Veen, E. M., Brentnall, A. R., Byers, H., Harkness, E. F., Astley, S. M., Sampson, S., Howell, A., Newman, W. G., Cuzick, J., & Evans, D. (2018). Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. JAMA oncology, 4(4), 476–482. https://doi.org/10.1001/jamaoncol.2017.4881

[34] Hüsing, A., Canzian, F., Beckmann, L., Garcia-Closas, M., Diver, W. R., Thun, M. J., Berg, C. D., Hoover, R. N., Ziegler, R. G., Figueroa, J. D., Isaacs, C., Olsen, A., Viallon, V., Boeing, H., Masala, G., Trichopoulos, D., Peeters, P. H., Lund, E.,

Ardanaz, E., Khaw, K. T et al (2012). Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status. Journal of medical genetics, 49(9), 601–608. https://doi.org/10.1136/jmedgenet-2011-100716

[35] Yanes, T., Young, M. A., Meiser, B., & James, P. A. (2020). Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field. Breast cancer research : BCR, 22(1), 21. https://doi.org/10.1186/s13058-020-01260-3

[36] Tice, J. A., Cummings, S. R., Smith-Bindman, R., Ichikawa, L., Barlow, W. E., & Kerlikowske, K. (2008). Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Annals of internal medicine, 148(5), 337–347. https://doi.org/10.7326/0003-4819-148-5-200803040-00004

[37] Tice, J. A., Miglioretti, D. L., Li, C. S., Vachon, C. M., Gard, C. C., & Kerlikowske, K. (2015). Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 33(28), 3137–3143. https://doi.org/10.1200/JCO.2015.60.8869

[38] Collaborative Group on Hormonal Factors in Breast Cancer (2012). Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. The Lancet. Oncology, 13(11), 1141–1151. https://doi.org/10.1016/S1470-2045(12)70425-4

[39] J. M., Castaño-Vinyals, G., Martín-Sanchez, V., Gómez-Acebo, I., Guevara, M., Tardón, A., Pérez-Gómez, B., Amiano, P., Moreno, V., Molina, A. J., Alonso-Molero, J., Moreno-Iribas, C., Kogevinas, M., Pollán, M., & Llorca, J. (2018). Reproductive risk factors in breast cancer and genetic hormonal pathways: a gene-environment interaction in the MCC-Spain project. BMC cancer, 18(1), 280. https://doi.org/10.1186/s12885-018-4182-3

[40] Zhang, X., Rice, M., Tworoger, S. S., Rosner, B. A., Eliassen, A. H., Tamimi, R. M., Joshi, A. D., Lindstrom, S., Qian, J., Colditz, G. A., Willett, W. C., Kraft, P., & Hankinson, S. E. (2018). Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. PLoS medicine, 15(9), e1002644. https://doi.org/10.1371/journal.pmed.1002644

Figures

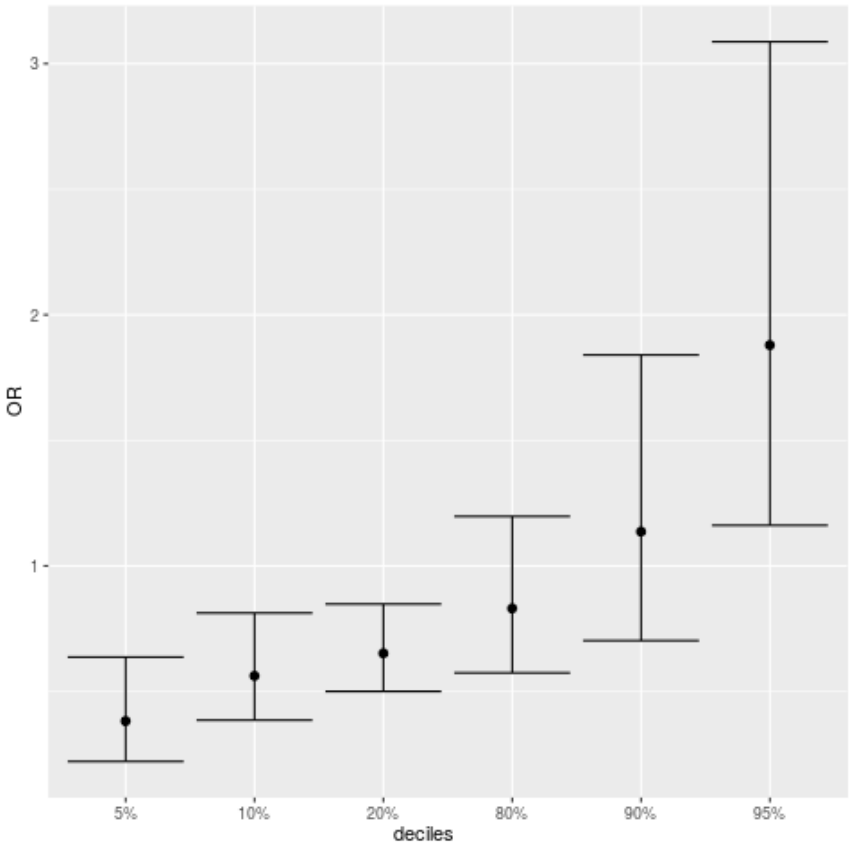

Figure 1

Association between the polygenic risk score (PRS92) and breast cancer risk in the Spanish population. The x-axis shows the different deciles and the y-axis the OR using the 40-60% range as reference.

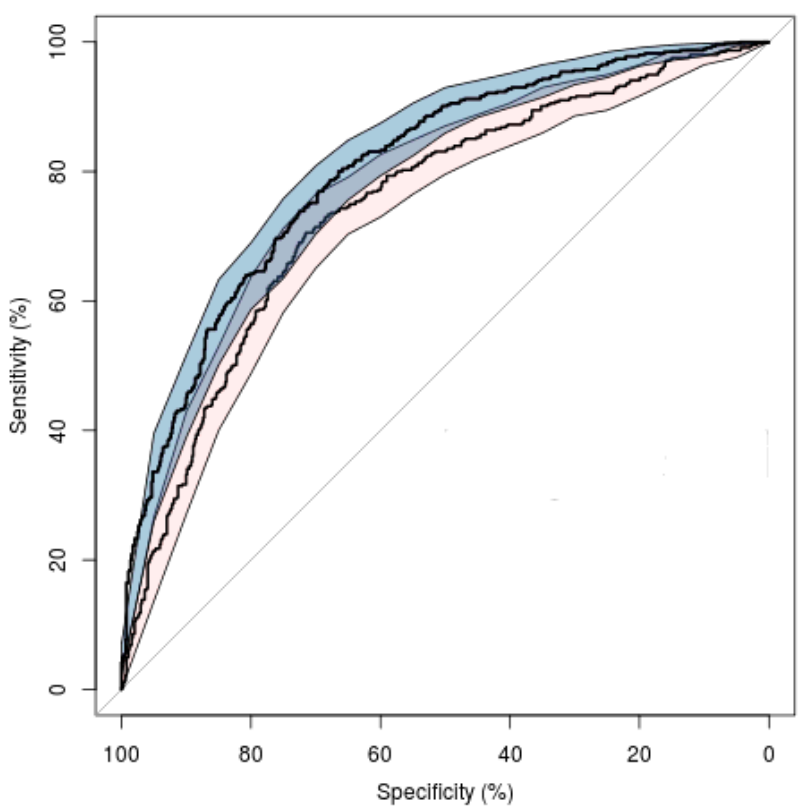

Figure 2

AUC-ROC of the multivariable model with and without interaction terms (blue and pink, respectively). The 95% confidence interval was evaluated using a bootstrap strategy.

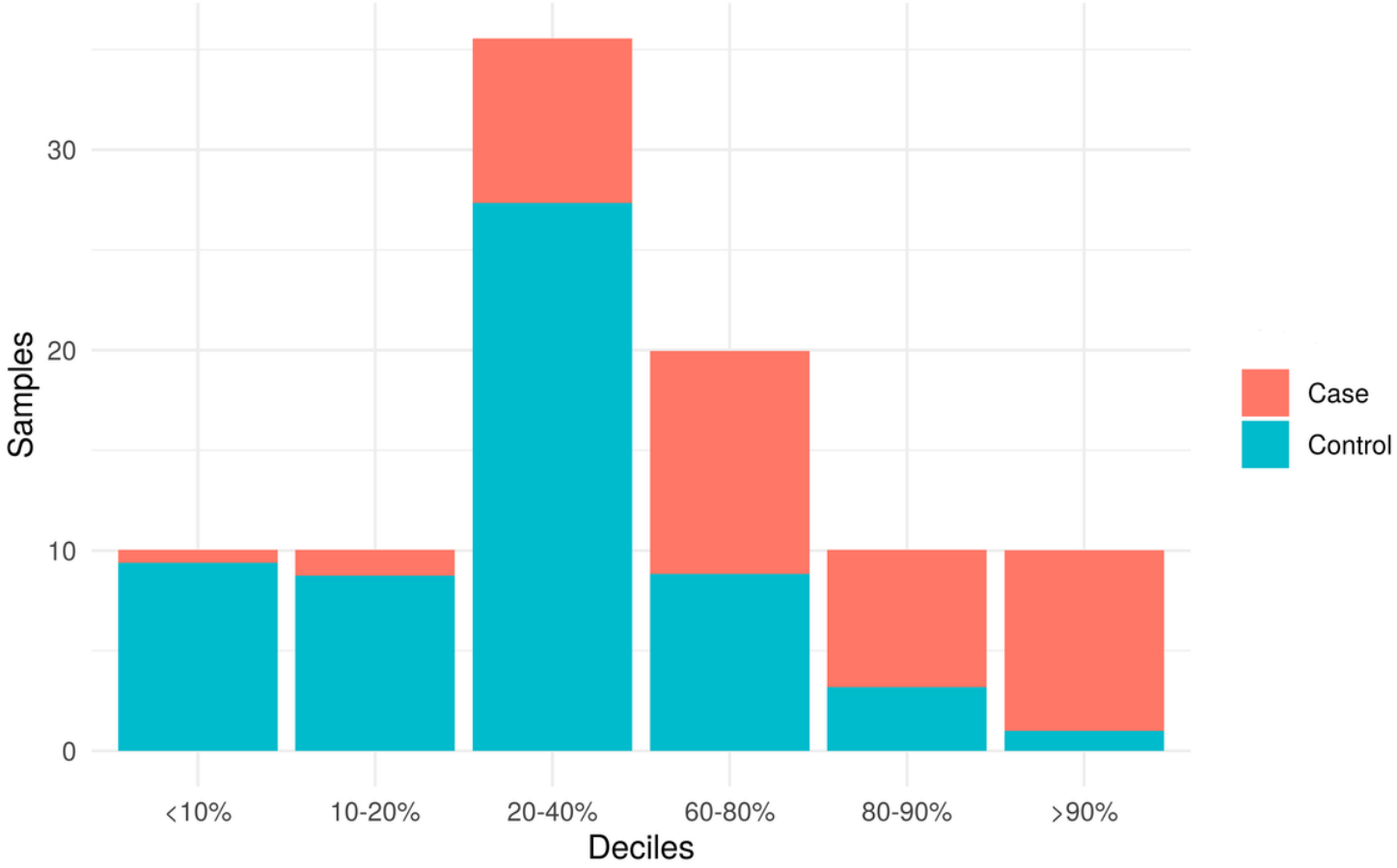

Figure 3

Case and control distribution using the multivariable model with interactions, and divided into deciles. The 40-60% range was selected as reference.

Supplementary Files

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

- Formula.pdf