

Editorial

Polygenic risk score as a key factor in cardiovascular clinical prediction models



La puntuación de riesgo poligénico como factor clave en los modelos de predicción clínica cardiovascular

María Brion,^{a,b,c,*} José Ramón González-Juanatey,^b and Ángel Carracedo^c^a Grupo de Xenética Cardiovascular, Instituto de Investigación Sanitaria de Santiago, Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, A Coruña, Spain^b Servicio de Cardiología, Complejo Hospitalario Universitario de Santiago, Instituto de Investigación Sanitaria de Santiago, CIBERCV, Santiago de Compostela, A Coruña, Spain^c Grupo de Medicina Xenómica, Universidad de Santiago de Compostela, Fundación Pública Galega de Medicina Xenómica, CIBERER, Santiago de Compostela, A Coruña, Spain

Article history:

Available online 17 February 2020

In clinical practice in the context of early disease detection and prevention of common adult-onset conditions, one of the most common problems is classification or decision-making, which is carried out through diagnostic and/or prognostic tests. When seeking their optimization, it is fundamental to be aware of their accuracy and precision.

Since the completion of the Human Genome Project, the combination of large-scale genome variation projects, such as the HapMap and 1000 Genomes projects, together with low-cost robust genotyping platforms and the rapid advance of DNA sequencing technologies, has enabled genome-wide association studies (GWAS) in large cohorts and exome- and genome-wide sequencing studies. Consequently, there has been an exponential increase in the abundance of individual-specific genotype data, leading to the era of personalized medicine or precision genomics-based medicine.¹

Historically, genetic diseases were classified into those with a Mendelian or simple inheritance caused by genetic variations with a large effect and those with a complex inheritance caused by the sum of genetic variations with a reduced effect. However, currently, each individual's overall risk of developing a common disease is probably marked by a combination of common low-risk genetic variants and rare high-risk genetic variants.²

GWAS have focused on identifying disease- or trait-associated genetic variants (typically single nucleotide polymorphisms [SNPs]), which are common in a given population (eg, minor allele frequency > 1%). To date, GWAS have identified thousands of loci that are associated with several complex human traits and diseases, including cardiovascular diseases.³ Notably, many of the loci previously associated with these complex human diseases are highlighted by multiple low-risk SNPs.⁴

These data have provided numerous insights into the genes and pathways that cause disease, but more recently there has been

increasing interest in the use of these data for disease risk prediction.^{5,6} In the last decade, genomics-based precision medicine has consistently emerged to provide effective and tailored health care for patients, depending on their genetic background. The inclusion of genetic risk scores (GRS), including disease or phenotype associated SNPs, into risk modelling has improved the accuracy of individual disease prediction,⁷ as reported in an original article published by Rincón et al. in *Revista Española de Cardiología*.⁸

The main focus of the development of genetic risk models is to achieve accurate predictive power for recognizing at-risk individuals (figure 1). Most commonly these models are calculated as a weighted sum of the number of risk alleles carried by an individual, where the risk alleles and their effect sizes are defined by previous GWAS.⁶ Therefore, the accuracy of a GRS is marked by the efficiency of previous GWAS studies in finding genetic variants associated with common diseases. In other words, the sounder the foundations of the building—in our case the genetic associations described in the GWAS studies—the more resistant our construction will be, ie, the more accurate our risk prediction estimate will be.

Predictive performance is typically evaluated by receiver operating characteristic (ROC) curves, in which the sensitivity and specificity of the predictions are ranked at various cutoff values. In the simplistic case, in which the development of any condition or disease is to be predicted, sensitivity is given as the fraction of the true-positive ratio among the total number of patients with the disease. Of note, a true positive is any patient who has the disease and has a positive result in the clinical prediction model. Therefore, the true-positive ratio is the probability of correctly classifying a patient. The specificity is the true-negative ratio among the total number of patients without the disease. Specificity is the probability of correctly classifying a healthy individual, that is to say, the probability that a healthy person will have a negative result. A ROC curve is a 2-dimensional graph in which the true-positive rate (sensitivity) is represented on the vertical axis, and the false-positive rate (1-specificity) on the horizontal axis. Therefore, a ROC graph of a prediction model represents the relative equilibrium between true positives and false negatives. The area under the ROC curve is the probability of

SEE RELATED CONTENT:

<https://doi.org/10.1016/j.rec.2019.08.006>

* Corresponding author: Complejo Hospitalario Universitario de Santiago de Compostela, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Laboratorio 1, Travesía da Choupana s/n, 15706 Santiago de Compostela, A Coruña, Spain.

E-mail address: maria.brion@usc.es (M. Brion).

<https://doi.org/10.1016/j.rec.2020.01.003>

1885-5857/© 2020 Sociedad Española de Cardiología. Published by Elsevier España, S.L.U. All rights reserved.

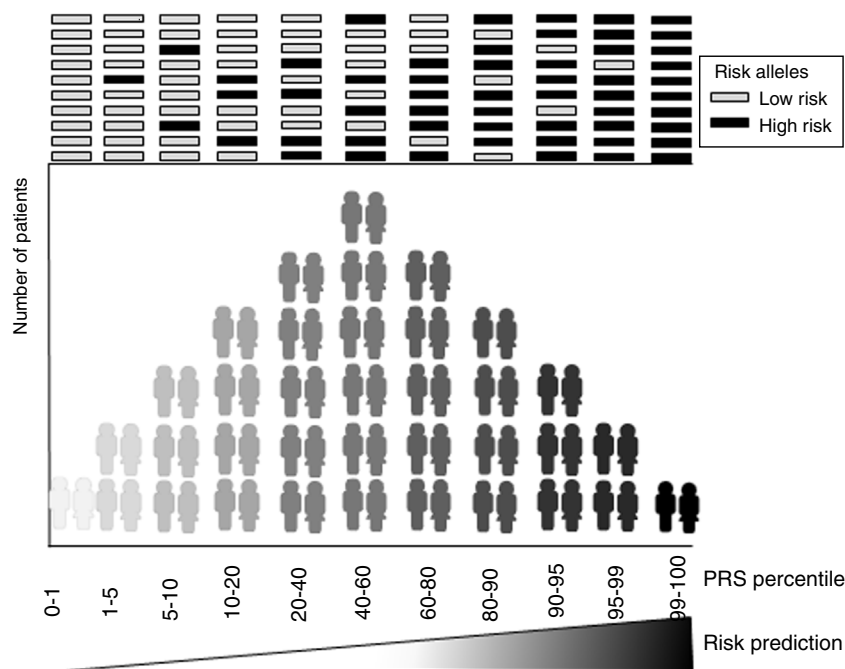


Figure 1. Risk score distribution. Distribution of the genetic risk ranges in a population according to the accumulation of risk alleles. PRS, polygenic risk score.

the examined model correctly identifying a case out of a randomly chosen pair of case and control samples. Area under ROC curve results range from 0.5 (ie, random) to 1 (ie, 100% accuracy).

GRS or polygenic risk scores (PRS), as many authors now call them, rather than predict the presence or absence of a disease, aim to classify the population into different risk levels (figure 1). The threshold for considering a positive GRS depends on the balance between the risk value marked by their own cutoff values in combination with other risk factors and the benefits associated with a possible therapeutic or lifestyle intervention.

It is currently believed that the genetics of nonfamilial forms of the most common adult-onset heart diseases are mainly linked to a combination of common variants with small effect sizes distributed throughout the genome and rare variants of moderate effect in genes known to cause familial disease. Evidence of this has been described in recent comprehensive genomic studies, such as an extensive GWAS coronary artery disease study⁹ and a large-scale sequencing study of type 2 diabetes mellitus.¹⁰ Therefore, the effect of each of these common variants on an individual will be too small to predict risk, but the combination of many of these common variants can be used to predict risk efficiently, especially if risk is predicted in combination with classic risk factors, such as clinical risk factors or certain environmental exposures.

One of the first publications on the implementation of GRS in cardiovascular diseases was the study by Morrison et al.,¹¹ in which the use of an 11-polymorphism score for predicting coronary heart disease risk did not improve the predictive capacity of classic risk factors. Since then, many PRS have been published and validated and are especially effective in groups of patients with highly specific phenotype. Some examples are coronary artery disease PRS aiming to individualize the decision to initiate lifetime statin therapy,¹² or PRS to improve the prediction capacity of patients classified as being at intermediate risk of cardiovascular heart disease according to the Framingham scale.¹³

GRS applied to young adults to predict recurrent events after myocardial infarction, as described by Rincón et al.,⁸ should be validated in a more extensive sample, since positive results are mainly seen in young patients without diabetes, but the observation is based on a small number of patients. Studies like

this one open doors to the implementation of PRS in clinical prediction models but they must always be validated and based on extensive data.

Last but not least, we must not forget the uncertainty in the estimation of the effect size associated with each common variant included in a genetic score, when the PRS is used to estimate the risk in other populations beyond the population studied in the GWAS. Since most of the GWAS were executed in European ancestry populations and genetic diversity among populations with different ancestry is well known, we must take special care when extending the applicability of PRS to all populations worldwide. Estimates are not transferable between populations, and ultimately the PRS is applied to an individual patient with a given geographical origin and with a characteristic genetic load, but with the same rights to health care.¹⁴

Although the number of studies with polygenic risk estimates has grown exponentially in the last 5 years, large-scale studies should be carried out to demonstrate the usefulness of polygenic risk estimation, not only in the cardiovascular field but also in other areas of human health. In this regard, the European action of the One Million Genomes Initiative, aiming to have this number of genomes linked to clinical data sequenced for 2022, with Spain a Signatory Member State, is perhaps the most promising project.¹⁵ Whole-genome data at this scale have the potential to make rapid progress in precision medicine and risk prediction estimates.

FUNDING

This work was partially supported by *Plan Estatal de I+D+i 2013-2016, Subdirección General de Evaluación y Fomento de la Investigación (ISCIII-SGEFI)* from *Instituto de Salud Carlos III (ISCIII)* and *Fondo Europeo de Desarrollo Regional (FEDER)* (grant numbers PI16/00903, CB16/11/00226, CB06/07/0088).

CONFLICTS OF INTEREST

None declared.

REFERENCES

1. Laksman Z, Detsky AS. Personalized medicine: understanding probabilities and managing expectations. *J Gen Intern Med.* 2011;26:204–206.
2. Katsanis N. The continuum of causality in human genetic disorders. *Genome Biol.* 2019;17:233.
3. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47:D1005–D1012.
4. Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet.* 2017;101:5–22.
5. Abraham G, Inouye M. Genomic risk prediction of complex human disease and its clinical application. *Curr Opin Genet Dev.* 2015;33:10–16.
6. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet.* 2018;19:581–590.
7. Wang X, Strizich G, Hu Y, Wang T, Kaplan RC, Qi Q. Genetic markers of type 2 diabetes: progress in genome-wide association studies and clinical application for risk prediction. *J Diabetes.* 2016;8:24–35.
8. Rincón LM, Sanmartín M, Alonso GL, et al. A genetic risk score predicts recurrent events after myocardial infarction in young adults. *Rev Esp Cardiol.* 2020;73:623–631.
9. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47:1121–1130.
10. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature.* 2019;536:41–47.
11. Morrison AC, Bare LA, Chambless LE, et al. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol.* 2007;166:28–35.
12. Mega JL, Stitzel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: An analysis of primary and secondary prevention trials. *Lancet.* 2015;385:2264–2271.
13. Iribarren C, Lu M, Jorgenson E, et al. Clinical Utility of Multimarker Genetic Risk Scores for Prediction of Incident Coronary Heart Disease: A Cohort Study Among Over 51 000 Individuals of European Ancestry. *Circ Cardiovasc Genet.* 2016;6:531–540.
14. Martin AR, Gignoux CR, Walters RK, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet.* 2017;100:635–649.
15. Saunders G, Baudis M, Becker R, et al. Leveraging European infrastructures to access 1 million human genomes by 2022. *Nat Rev Genet.* 2019;20:702.