Polygenic risk scores for complex diseases: where are we now?

BACKGROUND

Polygenic risk scores (PRS), commonly referred to as genetic or genomic risk scores, aggregate the effects of multiple genetic variants into a single composite estimate of genetic risk. PRS scores are typically used to predict the risk of developing a disease or to explain the phenotypic variation, and are derived from the effect sizes observed in large-scale genome-wide association studies (GWAS). Unlike rare monogenic diseases such as cystic fibrosis, which are attributable to genetic variants in single genes, with large effects on disease status, common complex diseases such as type 2 diabetes mellitus (T2DM) are polygenic, with risk contributed by a panel of genetic variants present throughout the genome. The concept of integrating information from these multiple genetic variants into a single metric of genetic risk was initially proposed in the shape of genetic risk scores, which generally limited the score to include single-nucleotide polymorphisms (SNPs) that were common and reached genome-wide significance in the initial GWASs. In contrast, PRS incorporates information from a much larger set of genetic variants, typically hundreds of thousands, including SNPs below the threshold for genome-wide statistical significance, and often with much more modest effect sizes.[1] Indeed, recent findings have pointed to how polygenic background could also increase the accuracy of risk estimation for individuals with monogenic risk variant in conditions such as familial hypercholesterolaemia, hereditary breast and ovarian cancer, and Lynch syndrome.[2]

DEVELOPMENT IN PRS APPROACHES

Initial PRS approaches were constructed using linear or logistic regression models to quantify individual SNP effect sizes. However, recent work has demonstrated that in the presence of linkage disequilibrium (LD), prediction accuracy of the commonly used approach of LD pruning/clumping followed by *P* value thresholding is inadequate.[3] More recent PRS approaches have used shrinkage and Bayesian methods such as *lassosum*, LDpred and PRS-CS.[3-5] In particular, Khera *et al*.[6] demonstrated that by taking a genome-wide PRS approach with LDpred, they were able to identify 1.5%–8.0% of the population at greater than three-fold increased risk of five common diseases, including coronary artery disease (CAD), atrial fibrillation, T2DM, inflammatory bowel disease and breast cancer. This was especially striking for CAD, where the observed prevalence of 8% is 20-fold higher than that of the carrier frequency of rare monogenic mutations conferring comparable risk. Indeed, these PRS provide predictive utility independently and additively to conventional clinical risk scores.[7,8] Having a high PRS contributed 21%–38% higher lifetime risk and 4–9 years earlier disease onset compared to an average PRS across common diseases. In fact, 13% of early-onset coronary heart disease cases were predicted only with the addition of PRS in the assessment model.[7] Work in UK Biobank also demonstrated in parallel that inclusion of PRS, in addition to traditional risk factors, increased approximately 7% in the number of events prevented.[8]

ROLE OF ETHNICITY IN PRS

Recent evidence has demonstrated the importance of using ethnic-appropriate PRS in disease risk prediction. Martin *et al*.[9] showed that PRS based on European-derived summary statistics had substantially lower accuracy when applied to non-European populations across 17 anthropometric and blood panel traits; accuracy was 1.6–4.5-fold lower on average in Hispanic/Latino Americans, South Asians, East Asians and Africans, compared to Europeans. This is corroborated by our own data (under review), in which we show that PRS for T2DM prediction in South Asians are significantly more accurate when based on South Asian rather than European association test results. These observations are particularly relevant to Asian populations such as Singapore, where we are strongly dependent on European-derived summary statistics for PRS, given that almost 80% of all current GWAS participants are of European ancestry.[9] The limited availability of genomic data for Asian populations underpins the current national effort in performing whole genome sequencing in a cohort of 100,000 individuals (SG100K) as part of Singapore's National Precision Medicine programme. This will allow us to assess the impact of combining an individual's PRS, lifestyle information and clinical information on chronic disease risk prediction and the associated implementation of early intervention, through lifestyle modification and/or early pharmacological intervention. Given the multi-ethnic make-up of the Singapore population, these population-specific PRS will prove invaluable in applying PRS to the greatest benefit in Singapore.

CHALLENGES AND THE FUTURE OF PRS UTILITY

Despite the potential benefits that clinical usage of PRS may bring, there remain many challenges and critical considerations in the journey to implementation. One concern is the quality control surrounding PRS generation, and where it should sit compared to typical clinical genetic testing, which is performed in Clinical Laboratory Improvement Amendments

(CLIA)-certified (or equivalent) laboratories and includes genetic counselling sessions. It is therefore essential to ensure education and training of healthcare professionals, as well as the general public, in the interpretation and understanding of these novel scores. Funding mechanisms should be in place to establish who should be responsible for the cost of genotyping and PRS generation, how best to handle any incidental findings, and whether PRS could be shared between primary health institutions and private practitioners. Public trust must be built, as PRS could be perceived by some as discriminating or stigmatising, and there may be concerns about its impact on one's insurance policies. There are also concerns around equal utility, and efforts such as SG100K, which aim to increase the genetic diversity of participants in GWAS, will help to improve utility for all groups, especially for underrepresented Asian communities. To ensure long-term successful implementation of PRS in routine clinical care, it is critical to demonstrate health and economic benefits. Public health and economic benefits will differ greatly depending on the actual use case for PRS, and considering possible clinical actions such as cost of intervention and screening strategies.[10] A significant health economics benefit could be achieved by the generation of multiple concurrent PRS for multiple diseases based on a single genotyping, and leveraging on the panel to optimise treatment and screening strategies across endpoints.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Marie Loh[1,2,3], MSc, PhD, John Campbell Chambers[1,2], FRCP, PhD

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, [2]Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom, [3]National Skin Centre, Singapore

Correspondence: Prof Marie Loh,
Lee Kong Chian School of Medicine, Nanyang Technological University, Clinical Sciences Building, 11 Mandalay Road, 308232, Singapore.
E-mail: marie_loh@ntu.edu.sg

Received: 16 Oct 2021 **Accepted:** 06 Nov 2021 **Published:** 19 Jan 2023

REFERENCES

1. Igo RP Jr, Kinzy TG, Cooke Bailey JN. Genetic risk scores. Curr Protoc Hum Genet 2019;104:e95.
2. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, *et al*. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nat Commun 2020;11:3635.
3. Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, *et al*. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am J Hum Genet 2015;97:576-92.
4. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genet Epidemiol 2017;41:469-80.
5. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun 2019;10:1776.
6. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, *et al*. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219-24.
7. Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, *et al*. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat Med 2020;26:549-57.
8. Sun L, Pennells L, Kaptoge S, Nelson CP, Ritchie SC, Abraham G, *et al*. Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. PLoS Med 2021;18:e1003498.
9. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019;51:584-91.
10. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet 2019;28:R133-42.

Access this article online

Quick Response Code:

Website: https://journals.lww.com/SMJ

DOI: 10.4103/singaporemedj.SMJ-2021-388

How to cite this article: Loh M, Chambers JC. Polygenic risk scores for complex diseases: Where are we now? Singapore Med J 2023;64:88-9.