

Will genetic data significantly change cardiovascular risk prediction in daily practice?

William J Young^{1,2}, Julia Ramírez^{1,3}, Stefan van Duijvenboden^{1,3}, Andrew Tinker¹, Pier D Lambiase^{2,3}, Patricia B Munroe¹, Michele Orini^{2,3}

¹Queen Mary University of London, London, United Kingdom

²Barts Heart Centre, St Bartholomew's Hospital, London, United Kingdom

³University College London, London, United Kingdom

Abstract

Precision medicine has been heralded as an opportunity to improve risk prediction, driven significantly by an increasing availability of genetic data. Genetic testing for rare mutations linked with Mendelian monogenic syndromes is available in specialised clinics. For complex diseases however, aggregation of common and low frequency variants into a “polygenic risk score” (PRS) is necessary due to their small individual effect sizes. PRSs for coronary artery disease (CAD), hypertension and atrial fibrillation have shown some modest success at a population level. However, scepticism remains whether the genetic effects in CV disease are sufficient to have meaningful clinical impact. This review explores recent efforts to utilise genomic data for risk prediction using CAD as an example.

1. Introduction

Despite significant advances in cardiovascular (CV) medicine over the last three decades, risk prediction and prevention of disease remains a challenge [1]. Precision medicine, an approach which seeks to optimise clinical care utilising variability at an individual level has been heralded as an opportunity to overcome this problem [2]. One component of this could be the use of genetic data which has become increasingly available for research as the reduction in cost of genotyping continues to fall, and could be used in the clinical environment [2]. As genotype is determined at birth, exploiting this information has potential to determine risk at an early stage before the onset of sub-clinical disease. Genetic testing for rare mutations linked with Mendelian monogenic syndromes is available in specialised clinics with appropriate genetic counselling and can guide clinical management. For instance, in Long QT syndrome, genotype may influence response to β -adrenergic blockade and incidence of sudden cardiac death [3]. However, many cardiovascular diseases with a major

public health burden are highly polygenic. Risk prediction for such diseases presents unique challenges including the construction of polygenic risk scores (PRS) using genetic variants associated with the disease which individually have a small effect on disease risk [4].

2. Construction of a PRS

A PRS is usually calculated by summing risk alleles which are weighted by effect sizes derived from genomic wide association study (GWAS), where the association of genetic variants with a phenotype or a surrogate marker of interest has been tested [5,6] Regression models can then be used to test ability of the PRS to predict clinical events. This method of transforming genetic information into a single number that quantifies a patient's propensity for developing the disease, poses challenges for which there is at present no standardised approach. For example, the selection of variants and weighting of effect sizes, a critical aspect of PRS calculation, is variable [7]. Different P-value thresholds, quality control metrics such as imputation quality, and the effect of underlying population stratification, can significantly influence the resulting PRS [7]. Various software, such as PLINK, LDpred and PRSice, have capability for selecting variants and constructing a PRS, but each employ different methods [8-10]. This reflects the challenges faced with optimising PRS for risk prediction in common complex diseases. Finally, causal variants and their effect sizes may differ according to population ancestry [11]. This needs to be considered when selecting variants for testing in the target population.

3. PRS for risk stratification of disease in the general population

Despite the above challenges, PRSs for coronary artery disease (CAD), hypertension and atrial fibrillation amongst others have had some modest success at a population level [12-14]. Initial PRSs for CAD were

significantly under-powered due to the relative smaller sample size of the original GWAS and thus fewer variants were available for testing [15]. For example, a PRS constructed using 13 CAD variants was associated with an increased risk of CAD (hazard ratio (HR) of 1.66 [95% CI: 1.35 to 2.04]) when comparing the top versus bottom quintiles in the PRS distribution [15]. Although this PRS translated into a substantial spread of risk in the population it did not outperform conventional risk factors. Improvement in the density of reference panels for genotype imputation and an increase in the sample size of GWAS, has led to an increase in the number of significant variants identified [5, 16]. In turn, a larger number of potential variants are available for generating a PRS. Inouye et al constructed a PRS for CAD with more than 1.7 million variants, representing 26.4% of the heritability of CAD (estimated heritability: 40-60%) [14]. This yielded a PRS for the prediction of CAD, with a higher c-index (or area under the ROC curve) than individual conventional risk factors including family history, smoking status, diabetes, hypertension and hypercholesterolaemia ($C = 0.603$; 95% CI: 0.615 to 0.631), with individuals in the top 20% of the PRS distribution being at more than 4-fold risk than those in the bottom 20%. Despite this, when combining the PRS with all conventional risk factors, only a small gain in discrimination was observed ($C=0.670$ for all conventional risk factors combined vs $C=0.696$ after integrating the PRS). Although showing that inclusion of genetic variants to clinical based scores independently contribute to predicting risk of CAD, this study was limited by the use of self-reported information for risk factors such as lipids. CV risk scores used in clinical practice commonly utilise biochemistry data, such as serum cholesterol/high-density lipoprotein, which improve risk stratification such as QRISK3 [17]. Thus, to effectively draw comparisons with tools currently used in CV medicine and understand the role of genotype in clinical practice, careful consideration needs to be made for the choice and measurement of variables included in models. Further study is also necessary to understand the relationship of some standard risk factors with genetic scores and how to best integrate both. For example, a family history of CAD is considered an important risk-factor and likely incorporates both genetic and environmental factors. Some evidence suggests a PRS may provide additive prognostic value to family history when included in the prediction of coronary events [18, 19].

The prospect of being able to screen at a population level for cardiovascular disease is enticing, as it offers potential for early intervention including lifestyle advice, targeted medical therapy and may guide frequency of clinical monitoring for disease. A PRS will not however, obviate the need for established clinical markers as non-modifiable risk factors, such as sex and age, along with environmental predictors, will continue to have an important role in risk prediction. Results from PRS

analyses therefore need to be evaluated in the correct context. The limited gains in predictive ability at a population level are perhaps unsurprising at this stage, given the need for further optimisation of existing methods for constructing a PRS. However even beyond CV disease, PRS continues to offer potential for the prediction of disease [20].

4. Role of PRS in patient sub-groups

A PRS may have a role beyond risk prediction at a general population level. Ellis et al explored the predictive value of a PRS in a cohort of patients with Familial Hypercholesterolaemia (FH) and compared with a history of CAD [21]. This study, although limited by the lack of prospective follow up, showed a significantly higher odds ratio for a history of CAD with increasing PRS. This association was stronger in patients with a recognised mutation associated with FH in a multivariate analysis including clinical risk factors. This study in a real-world clinical setting, highlights the potential application of PRS in high risk cohorts to guide early intervention. It also suggests there may be a role in utilising common variation in patients with monogenic conditions, which is also of interest in inherited channelopathies [22]. Other high risk patient groups worthy of separate testing of a CAD PRS include those with a history diabetes, chronic kidney disease and hypertension.

5. PRS in the prediction of response to treatment

Response to medical therapy is governed by several mechanisms, including the pharmacodynamic and pharmacokinetic properties of a drug. Such mechanisms are also influenced by genetic variation. Polymorphisms in Cytochrome P450 (CYP450) enzymes are well recognised to alter metabolism and excretion of drugs, such as clopidogrel, an anti-platelet agent used in the management of acute coronary syndrome (ACS) and stroke [23]. A pharmacogenetic polygenic risk score using 31 candidate CYP450 polymorphisms was associated with an increase in cardiovascular events and cardiovascular-related death with increasing number of effect alleles, in a cohort of patients taking clopidogrel [24]. Additionally, a post-trial ad-hoc analysis of Alirocumab, an antibody that blocks PCSK9 (proprotein convertase subtilisin/kexin type 9) compared response to therapy with a CAD PRS in patients with ACS and elevated serum lipid levels [25]. This analysis identified an absolute reduction of 6% and relative risk reduction of 37% (HR 0.63 (95%CI: 0.40 – 0.86) for the incidence of CAD related events in patients in the top 10% of PRS distribution taking Alirocumab compared with placebo. This was compared to a smaller risk reduction of 13% (HR 0.87) in patients in the lower 10%.

Similar findings have also been identified in a nested case-control study for Evacetrapib [26]. These studies, while preliminary, and thus should be approached with caution, identify potentially important roles for a PRS in the selection of medical therapy. A large proportion of ischaemic events following percutaneous coronary intervention, including stent thrombosis, occur while patients are still on anti-platelet agents such as Clopidogrel [27]. Such events are likely in-part due to individual variability in the response to drug treatment [28]. Therefore in this scenario, using personalised information such as genotype to predict response to a specific drug, could inform the choice of anti-platelet used, thus reducing stent related complications. Side-effects from medication have a significant impact on healthcare services, morbidity and mortality and mental health [29]. A US study reported over 6% of hospitalisation were due to adverse drug related events [30]. Targeted medical therapy aided by a PRS could reduce the number of patients exposed to side-effects when a reduction in risk of CV events is likely to be limited and aid the prioritisation of resources.

6. Limitations and future perspectives

A major current limitation for the successful implementation of PRS in clinical practice is that the overwhelming majority of published studies are from European-ancestry populations. A PRS may be more predictive in one ancestry compared with another, particularly when using variants from a GWAS from an ancestry different to the test population, as causal variants, allele frequencies or effect sizes may significantly vary [11]. A recent study evaluating a CAD PRS derived in South-Asian cohorts identified a significantly increased risk for CAD in individuals in the top 5% of the GRS distribution [31]. However, such studies are rare, often under-powered and require expansion going forward.

Beyond evaluating the predictive value of a PRS, education for patients and health-care professionals needs to be disseminated. A wider discussion is also necessary on the use of genetic data covering informed consent, data governance and the perception of what risk means to the patient, medical staff and the global community.

Scepticism remains whether the genetic effects in CV traits are large enough to have meaningful clinical impact. For CV disease, many conditions have a gap where a significant proportion of the heritability is yet to be explained, however this may be address with further improvements in association testing for example with whole genome sequencing.

The association between PRS and the trait being studied is usually assessed in terms of the relative risk of individuals at the two opposite extreme of the PRS distribution, with odds or hazard ratios typically < 5 . However, it has been demonstrated that a PRS, as is the case with any other risk factor, may need to be

characterised by a much higher odds or hazard ratio (>100) to show both high sensitivity and specificity to be used as an effective screening tool [34]. For instance, despite showing a hazard ratio of 4.17 (comparing extreme quintiles), the recent PRS for CAD developed by Inouye *et al* showed low discriminative power (area under the ROC curve ~ 0.62) [14]. According to a risk screening converter, this would be consistent with a sensitivity of just 13% for a specificity of 95% [35, 36]. This however may also reflect that risk using a PRS is not linearly distributed and individuals at the extreme ends of the PRS distribution will have a greater cumulative risk.

7. Conclusion

The link between the genome and clinical outcomes is complex and yet to be clearly understood however there has been some promise with early work utilising PRS for CV risk prediction and personalisation of clinical management. Continued improvement in the methods for constructing a PRS, identification of additional variants associated with disease and testing in non-European ancestry populations, may improve the discriminative power of the PRS.

Acknowledgments

WJY is supported by a Medical Research Council grant MR/R017468/1. JR acknowledges support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No.786833. PDL is supported by UCL/UCLH Biomedicine NIHR. We also acknowledge the Medical Research Council grant MR/N025083/1 and the NIHR Cardiovascular Biomedical Centre at Barts and The London, Queen Mary University of London

References

- [1] F. Farzadfar, "Cardiovascular disease risk prediction models: challenges and perspectives," *Lancet Glob Health*, vol. 7, no. 10, pp.e1288-e1289, 10 2019..
- [2] M. J. Joyner, "Precision Medicine, Cardiovascular Disease and Hunting Elephants," *Prog Cardiovasc Dis*, vol. 58, no. 6, pp.651-60, 2016 May-Jun 2016..
- [3] P. J. Schwartz and M. J. Ackerman, "The long QT syndrome: a transatlantic clinical approach to diagnosis and therapy," *Eur Heart J*, vol. 34, no. 40, pp.3109-16, Oct 2013.
- [4] M. Ala-Korpela and M. V. Holmes, "Polygenic risk scores and the prediction of common diseases," *Int J Epidemiol*, vol. 49, no. 1, pp.1-3, 02 2020.
- [5] M. Nikpay *et al.*, "A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease," *Nat Genet*, vol. 47, no. 10, pp.1121-1130, Oct 2015.
- [6] C. M. Lewis and E. Vassos, "Polygenic risk scores: from

- research tools to clinical instruments," *Genome Med*, vol. 12, no. 1, p.44, 05 2020.
- [7] A. C. J. W. Janssens, "Validity of polygenic risk scores: are we measuring what we think we are?," *Hum Mol Genet*, vol. 28, no. R2, pp.R143-R150, 11 2019.
- [8] S. Purcell *et al.*, "PLINK: a tool set for whole-genome association and population-based linkage analyses," *Am J Hum Genet*, vol. 81, no. 3, pp.559-75, Sep 2007.
- [9] S. W. Choi and P. F. O'Reilly, "PRSice-2: Polygenic Risk Score software for biobank-scale data," *Gigascience*, vol. 8, no. 7, 07 2019.
- [10] B. J. Vilhjálmsón *et al.*, "Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores," *Am J Hum Genet*, vol. 97, no. 4, pp.576-92, Oct 2015.
- [11] H. Mostafavi, A. Harpak, I. Agarwal, D. Conley, J. K. Pritchard, and M. Przeworski, "Variable prediction accuracy of polygenic scores within an ancestry group," *Elife*, vol. 9, Jan 2020.
- [12] S. H. Choi *et al.*, "Monogenic and Polygenic Contributions to Atrial Fibrillation Risk: Results From a National Biobank," *Circ Res*, vol. 126, no. 2, pp.200-209, 01 2020.
- [13] H. R. Warren *et al.*, "Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk," *Nat Genet*, vol. 49, no. 3, pp.403-415, Mar 2017.
- [14] M. Inouye *et al.*, "Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention," *J Am Coll Cardiol*, vol. 72, no. 16, pp.1883-1893, 10 2018.
- [15] S. Ripatti *et al.*, "A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses," *Lancet*, vol. 376, no. 9750, pp.1393-400, Oct 2010.
- [16] A. R. Wood *et al.*, "Imputation of variants from the 1000 Genomes Project modestly improves known associations and can identify low-frequency variant-phenotype associations undetected by HapMap based imputation," (in eng), *PLoS One*, vol. 8, no. 5, p.e64343, 2013.
- [17] National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification. www.nice.org.uk/guidance/cg181
- [18] P. Antiochos, P. Marques-Vidal, A. McDaid, G. Waeber, and P. Vollenweider, "Association between parental history and genetic risk scores for coronary heart disease prediction: The population-based CoLaus study," *Atherosclerosis*, vol. 244, pp.59-65, Jan 2016.
- [19] N. Timmerman *et al.*, "Family history and polygenic risk of cardiovascular disease: Independent factors associated with secondary cardiovascular events in patients undergoing carotid endarterectomy," *Atherosclerosis*, May 2020.
- [20] 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.
- [21] K. L. Ellis *et al.*, "A genetic risk score predicts coronary artery disease in familial hypercholesterolaemia: enhancing the precision of risk assessment," *Clin Genet*, vol. 97, no. 2, pp.257-263, 02 2020.
- [22] N. Lahrouchi *et al.*, "Transethnic Genome-Wide Association Study Provides Insights in the Genetic Architecture and Heritability of Long QT Syndrome," *Circulation*, vol. 142, no. 4, pp.324-338, Jul 2020.
- [23] D. Tousoulis *et al.*, "The role of the cytochrome P450 polymorphisms in clopidogrel efficacy and clinical utility," *Curr Med Chem*, vol. 18, no. 3, pp.427-38, 2011.
- [24] J. P. Lewis *et al.*, "Pharmacogenomic polygenic response score predicts ischaemic events and cardiovascular mortality in clopidogrel-treated patients," *Eur Heart J Cardiovasc Pharmacother*, vol. 6, no. 4, pp.203-210, Jul 2020.
- [25] A. Damask *et al.*, "Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial," *Circulation*, vol. 141, no. 8, pp.624-636, Feb 2020.
- [26] C. A. Emdin *et al.*, "Genome-Wide Polygenic Score and Cardiovascular Outcomes With Evacetrapib in Patients With High-Risk Vascular Disease: A Nested Case-Control Study," *Circ Genom Precis Med*, vol. 13, no. 1, 02 2020.
- [27] R. Mehran *et al.*, "Cessation of dual antiplatelet treatment and cardiac events after percutaneous coronary intervention (PARIS): 2 year results from a prospective observational study," *Lancet*, vol. 382, no. 9906, pp.1714-22, Nov 2013.
- [28] J. Torrado *et al.*, "Restenosis, Stent Thrombosis, and Bleeding Complications: Navigating Between Scylla and Charybdis," *J Am Coll Cardiol*, vol. 71, no. 15, pp.1676-1695, 04 2018.
- [29] J. Sultana, P. Cutroneo, and G. Trifirò, "Clinical and economic burden of adverse drug reactions," *J Pharmacol Pharmacother*, vol. 4, no. Suppl 1, pp.S73-7, Dec 2013.
- [30] D. R. Poudel, P. Acharya, S. Ghimire, R. Dhital, and R. Bharati, "Burden of hospitalizations related to adverse drug events in the USA: a retrospective analysis from large inpatient database," *Pharmacoepidemiol Drug Saf*, vol. 26, no. 6, pp.635-641, Jun 2017.
- [31] M. Wang *et al.*, "Validation of a Genome-Wide Polygenic Score for Coronary Artery Disease in South Asians," *J Am Coll Cardiol*, vol. 76, no. 6, pp.703-714, Aug 2020.
- [32] J. I. Rotter and H. J. Lin, "An Outbreak of Polygenic Scores for Coronary Artery Disease," *J Am Coll Cardiol*, vol. 75, no. 22, pp.2781-2784, Jun 2020.
- [33] A. V. Khera *et al.*, "Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations," *Nat Genet*, vol. 50, no. 9, pp.1219-1224, 09 2018.
- [34] N. J. Wald, A. K. Hackshaw, and C. D. Frost, "When can a risk factor be used as a worthwhile screening test?," *BMJ*, vol. 319, no. 7224, pp.1562-5, Dec 1999.
- [35] N. J. Wald and J. K. Morris, "Assessing risk factors as potential screening tests: a simple assessment tool," *Arch Intern Med*, vol. 171, no. 4, pp.286-91, Feb 2011.
- [36] N. J. Wald and R. Old, "The illusion of polygenic disease risk prediction," *Genet Med*, vol. 21, no. 8, pp.1705-1707, 08 2019.

Address for correspondence:

Name: William J Young

Full postal address: William Harvey Research Institute, Charterhouse Square, Queen Mary University of London, London, UK, EC1M 6EA.

E-mail address: w.young@qmul.ac.uk