Clinical relevance of genome-wide polygenic score may be less than claimed

David Curtis

d.curtis@ucl.ac.uk

00 44 7973 906 143

UCL Genetics Institute, UCL, Darwin Building, Gower Street, London WC1E 6BT.
Centre for Psychiatry, Barts and the London School of Medicine and Dentistry, Charterhouse Square, London EC1M 6BQ.

Kaywords

Genome-wide polygenic score, coronary artery disease, AUC.

Summary

A recent study claimed that genome-wide polygenic scores (GPSs) for five common diseases could identify individuals with risk equivalent to monogenic mutations. Receiver operator curve analyses were reported to have areas under the curve (AUCs) ranging from 0.63 for inflammatory bowel disease up to 0.81 for coronary artery disease (CAD). The GPS for CAD identified 8% of the population at threefold increased risk, which it was claimed was comparable to the excess risk from monogenic mutations.

Attempts were made to model the distribution of GPS for CAD to match the information provided. Models were based on the reported distribution of prevalence and GPS and were fitted to the reported results using linear approximations to the distributions and using simulations of a liability-threshold model.

It was impossible to produce a compatible model which produced an AUC as high as 0.81 and the most plausible estimate was that the true AUC was only 0.65. The reported distributions in cases and controls largely overlap so that they are not compatible with an AUC of 0.7 or more.

The true AUC of the GPS for these diseases is much lower than claimed. Furthermore, the literature robustly demonstrates that true CAD risk associated with monogenic mutations is much higher than the threefold increase which was claimed. Together, these findings cast doubt on the clinical utility of the GPS.

Introduction

A recent study has claimed that genome-wide polygenic scores (GPS) for five common diseases could identify individuals with risk equivalent to monogenic mutations (Khera et al. 2018). Using GPS to predict risk it was reported that the area under the receiver operating curve (AUC) was high for all five diseases tested, ranging from 0.63 for inflammatory bowel disease up to 0.81 for coronary artery disease (CAD). Figures were presented which provided some information about the

distribution of GPS for CAD in controls and cases and the prevalence of CAD across different centiles of the GPS. It was reported that for CAD the GPS could identify 8% of the population at greater than threefold increased risk, and this was contrasted with the 0.4% carrying a familial hypercholesterolemia (FH) mutation which, it was claimed, "confers an up to threefold increased risk for CAD".

Here I model the system based on the information presented in these figures and the accompanying text and show that the true AUCs cannot be as high as claimed. I also point out that the FH mutation has a much larger than threefold effect on risk, directly challenging the claim that the GPS identifies "individuals with risk equivalent to monogenic mutations".

Methods

The relationship between GPS and CAD risk was modelled in three different ways.

The first approach used a simple approximation to the graph of CAD prevalence (%) against GPS centile which had been presented in Figure 2c of the original report. The prevalence (%) was reported to be 0.8 for the lowest centile and 11.1 for the highest but the curve is markedly concave so two linear segments were used (with x=centile, y=prevalence), from (1,0.8) to (90,4) and then from (90,4) to (100,11.1). From these interpolated prevalences, the sensitivity and specificity were calculated for each GPS centile and these values were used to calculate the AUC.

The second approach used the GPS medians and intra-quartile ranges to model the sensitivity and specificity. The medians for controls and cases were reported as 49 and 69 respectively, and from Figure 2b of the original report it appears that the intra-quartile ranges are approximately 23-73 and 44-89. This meant that the distribution of specificity for controls could be modelled (with x=centile, y=specificity) by having four linear segments passing through the points (0,0; 23,0.25; 49,0.5; 73, 0.75; 100,1) and sensitivity for cases with four linear segments passing through the points (0,1; 44,0.75; 69,0.5; 89,0.25; 100,0). Again, the AUC was then calculated from these values.

The third approach was to simulate a liability threshold model with the liability as the sum of a normally distributed component tagged by the GPS and a normally distributed residual component. These two components were assumed to contribute additively to the liability and all subjects with a resultant liability exceeding a threshold were designated cases. 100,000 subjects were used in the simulation. The model is parameterised by the proportion of liability variance accounted for by the GPS, denoted P, and the population prevalence K. Trial and error was used to select values of P and K which would produce a graph of prevalence against GPS centile which approximated the original Figure 2c. The distribution of GPS in cases and controls was then examined, along with the specificity, sensitivity and AUC.

Results

The first two approaches yielded values for the AUC of only 0.66 and 0.64 respectively. For the second approach, values for P and K of 0.06 and 0.04 produced distributions which seemed to match

fairly well with those originally reported. The graph of prevalence against centile of PGS is shown in Figure 1 and it can be seen that this approximates to the distribution shown in Figure 2c of the original publication. The values chosen produce control and case medians of 51 and 71 with IQRs of 26-76 and 46-90, acceptably close to the originals. The AUC for this model is only 0.65. In general, the value for the AUC is not very sensitive to the parameter values chosen. It was possible to increase the AUC to 0.80 by increasing P to 0.17 and reducing K to 0.01. However with these values the medians were 51 and 88 and IQRs of 26-76 and 71-98, completely discrepant from those reported for the real data. No model could be found with an AUC of 0.7 or more which yielded values even approximately resembling those displayed in the original Figures 2b and 2c. There is a large overlap in IQRs and the medians of the centiles are only separated by 20 and these features imply that no compatible model can produce an AUC as high as 0.7.

Discussion

The original claim that the AUC for risk of CAD predicted by GPS is 0.81 appears to be incompatible with the information which was presented. The IQRs largely overlap and the medians are not far enough apart to produce a high AUC and the best estimate for the true value is around 0.65. The reason for this discrepancy cannot be determined without access to the original data and analyses. It was reported that the AUC was obtained in a logistic regression model incorporating GPS and covariates but this cannot explain the incompatibility between the claimed AUC and the GPS distribution presented. The discrepancy is important because the AUC is widely used as a rule of thumb as to whether a screening test is likely to be clinically relevant. The other diseases studied were reported to have similar risk profiles with smaller claimed AUCs than for CAD so it seems reasonable to assume that their true AUCs are also low. In general, a test with an AUC of 0.65 or less would not be expected to be used in clinical practice.

Another problem is the claim that the FH mutation "confers an up to threefold increased risk". The source cited to produce a this estimate of risk for monogenic mutations (OR around 3) was based on a sample of exome-sequenced subjects in a health insurance scheme with a median age of 61 (Abul-Husn et al. 2016). Such a sample would be expected to be significantly depleted for relevant mutations because subjects with them might have died or been removed from the scheme, leading to an under-estimate of their effect size. However, according to the consensus statement of the European Atherosclerosis Society the real effect size is much higher, "up to 13-fold increased risk", and "men and women with heterozygous FH typically develop CHD before age 55 and 60, respectively, while homozygotes typically develop CHD very early in life and if untreated die before age 20" (Nordestgaard et al. 2013). However, for subjects in the very highest centile of GPS for CAD, the prevalence of CAD was stated to be only 11.1%. Thus, the claim that the risk identified by GPS is "comparable" with the risk from monogenic mutations does not seem to be justified.

Given that the performance of the GPS appears to have been overstated and that the importance of monogenic mutations has been understated, it cannot be concluded that "genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations".

References

Abul-Husn, N.S. et al., 2016. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science (New York, N.Y.)*, 354(6319), p.aaf7000.

Khera, A. V. et al., 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9), pp.1219–1224.

Nordestgaard, B.G. et al., 2013. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: Consensus Statement of the European Atherosclerosis Society. *European Heart Journal*, 34(45), pp.3478–3490.

Figure 1 Graph of prevalence against PGS centile using a liability threshold model with P=0.06, K=0.04. The distribution is similar to that shown in Figure 2c of the original report.