



The illusion of polygenic disease risk prediction

Nicholas J. Wald, FRS ¹ and Robert Old, PhD¹

A problem at the interface of genomic medicine and medical screening is that genetic associations of etiological significance are often interpreted as having predictive significance. Genome-wide association studies (GWAS) have identified many thousands of associations between common DNA variants and hundreds of diseases and benign traits. This knowledge has generated many publications with the understandable expectation that it can be used to derive polygenic risk scores for predicting disease to identify those at sufficiently high risk to benefit from preventive

intervention. However, the expectation rests on the incorrect assumption that odds ratios derived from polygenic risk scores that are important etiologically are also directly useful in risk prediction and population screening.

Genetics in Medicine (2019) 21:1705–1707; <https://doi.org/10.1038/s41436-018-0418-5>

Two widely publicized recent papers (Khera et al.¹ and Inouye et al.²) illustrate the problem. These papers show associations between polygenic risk scores and a number of common disorders, including coronary artery disease (CAD). The results demonstrate the importance of genetic variation in the etiology of the disorders, but not the value of the risk score proposal in disease prediction (i.e., screening), contrary to what is suggested in these papers. The authors suggest that polygenic risk scores could be used to prompt preventive intervention among individuals with a high score but not among those with a low score. This, however, is based on a misconception that estimates of relative risk such as odds ratios or hazard ratios can directly and adequately assess the discriminatory value of polygenic risk scores as screening tests. For example, Khera et al. show that a CAD risk score that identifies 5% of people with the highest scores compared with people with the lowest risk scores had a CAD odds ratio of 3.34, which can create the impression of useful discrimination between CAD and non-CAD.¹ Similarly, Inouye et al. show that people with the highest 20% of risks using their proposed polygenic risk score algorithm have a hazard ratio of 4.17 for CAD compared with people in the bottom 20%.²

The problem, however, is that an odds ratio or hazard ratio does not directly indicate the discriminatory value of a screening test. To assess the discriminatory value, it is, whenever possible, necessary to specify the detection rate (sensitivity) and risk score cut-off for a given false-positive rate or the false-positive rate and risk score cut-off for a given detection rate.³ The detection rate is the proportion of

affected individuals with a positive score. The false-positive rate (1–specificity) is the proportion of unaffected individuals with a positive score. Affected individuals are those who develop the predicted disorder over a given period of time and unaffected individuals are those who remain free of the disorder over the same period.

The fact that a strong risk factor can be a poor screening test may seem counterintuitive. The paradox is largely explained by the fact that odds ratios or hazard ratios typically compare risks in the tails of a single risk distribution, but these ratios ignore the proportions of individuals who will or will not develop the disease that fall in the region between the tails of the distribution. The subject is discussed in detail in a previous publication,³ which explains how detection and false-positive rates can be calculated when only the odds ratio and the size of the centile groups are given.

Information on odds ratios can be converted into relevant measures of screening performance. This can be done using the published Risk Screening Converter,⁴ which is freely available on the Internet (<https://www.medicalscreeningsociety.com/rsc.asp>). The Risk Screening Converter shows that the Khera et al. polygenic risk score gives a CAD detection rate of 15% for a 5% false-positive rate, which means that the score would classify 5% of unaffected individuals as positive and would miss 85% of affected individuals. With the Inouye et al. score, the detection rate is 13% for a 5% false-positive rate. Altering the score cut-off level alters the detection rate and the false-positive rate, for example, yielding a 10% detection rate for a 3% false-positive rate using the Khera et al. score, or 8% using the Inouye et al.

¹Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London, UK.
Correspondence: Nicholas J. Wald (n.j.wald@qmul.ac.uk)

Submitted 26 October 2018; accepted: 17 December 2018

Published online: 12 January 2019

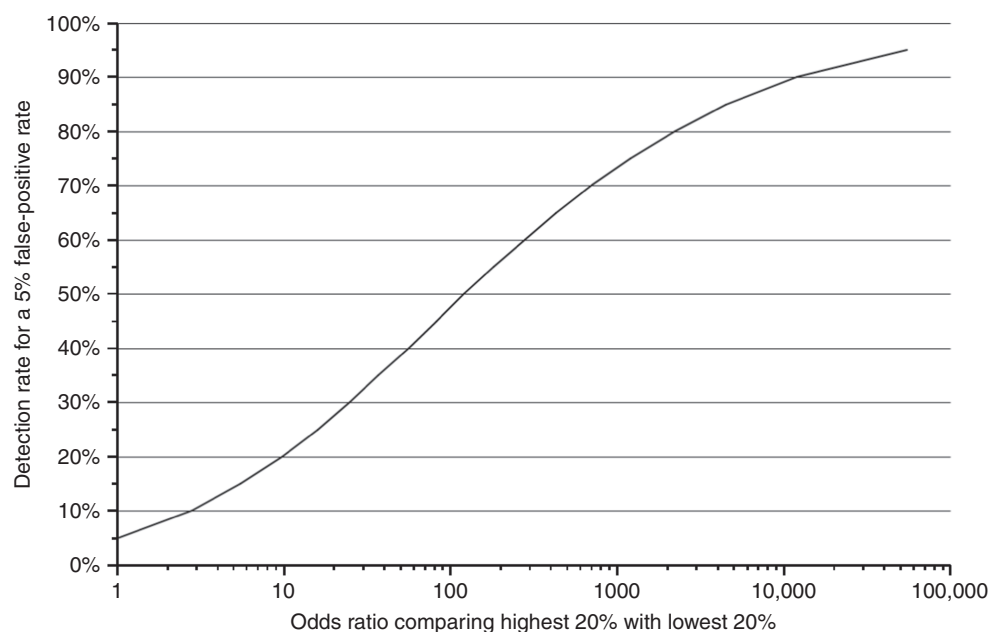


Fig. 1 Detection rate for a 5% false-positive rate according to the odds ratio of becoming affected among people in the highest compared with those in the lowest fifths of the distribution of a risk factor or risk score (adapted from refs. ³ and ⁴). The standard deviations of the separate distributions in affected and unaffected individuals are taken to be the same.

score. At a 10% false-positive rate, the detection rates are 25% and 22% respectively. Whatever the chosen cut-off, the screening performance is poor. Interested readers can use the Risk Screening Converter to evaluate other polygenic risk score studies, such as Schumacher et al.⁵ in predicting prostate cancer, quoting a relative risk of 5.71 in people with the highest 1% of risk compared with the population average.

Estimating odds ratios or hazard ratios is appropriate and customary in etiological studies but can be deceptive, and conceal the poor discriminatory power of predictive scores. Identifying about 15% of cases for a false-positive rate of 5% is poor discrimination and little better than identifying people at random. In such circumstances, if the proposed intervention is effective, inexpensive, and safe it would be better to offer the intervention without prior testing and save the cost of testing everyone. A very high odds ratio between the highest and lowest quintile groups (fifths) of the distribution of a risk factor or risk score is needed to be a useful screening test; even an odds ratio of 100 detects fewer than half (48%) of affected individuals for a 5% false-positive rate (see Fig. 1, which shows the relation between relative odds and detection rates for a 5% false-positive rate).^{3,4}

Some authorities⁶ recognize that polygenic risk scores are weak predictors of disease, but suggest that they could usefully be adopted in “risk stratification,” with the implication that specifying gradations of risk can overcome the problem.⁷ Risk stratification cannot, however, transform a weak predictor into a strong one. If a polygenic risk score is used in combination with one or more existing screening markers, the incremental gain in screening performance needs to be quantified by the increase in the detection rate for a given false-positive rate, or vice versa, and assessed in relation to the

extra cost. In exceptional circumstances, risk stratification may be warranted, for example, if screening leads directly to preventive intervention that is hazardous or costly (such as surgery following screening to prevent ruptured aortic aneurysm).

In summary, moderate relative risks (e.g., about 3-6) can have considerable significance in determining causes of disease. However, it is not well recognized that estimates of the relative risk between a disease marker and a disease have to be extremely high for the risk factor to merit consideration as a worthwhile screening test. To our knowledge, no genome-wide polygenic score meets this requirement, and none is likely to do so with polygenic scores that emerge in the future. It is important that the potential applications of genomic medicine are not compromised by raising unrealistic expectations in medical screening.

DISCLOSURE

The authors declare no conflicts of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Khera AV, Chaffin M, Aragam KG et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50:1219–1224. <https://doi.org/10.1038/s41588-018-0183-z>
2. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F et al. Genomic risk prediction of coronary artery disease in 280,000 adults: Implications for primary prevention. *J Am Coll Cardiol* 2018;72. <https://doi.org/10.1016/j.jacc.2018.07.079>

3. Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? *BMJ*. 1999;319:1562–1565.
4. Wald NJ, Morris JK. Assessing risk factors as potential screening tests: a simple assessment tool. *Arch Intern Med*. 2011;171:286–291.
5. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet*. 2018;50:928–936.
6. Khoury MJ, Janssens AC, Ransohoff DF. How can polygenic inheritance be used in population screening for common diseases? *Genet Med*. 2013;15:437–443.
7. Chowdhury S, Dent T, Pashayan N, et al. Incorporating genomics into breast and prostate cancer screening: assessing the implications. *Genet Med*. 2013;15:423–432.