bmjmedicine

Check for updates

Performance of polygenic risk scores in screening, prediction, and risk stratification: secondary analysis of data in the Polygenic Score Catalog

Aroon D Hingorani,[1,2,3,4] Jasmine Gratton [ORCID],[1,2] Chris Finan,[1,2,3,4] A Floriaan Schmidt,[1,3,4,5] Riyaz Patel,[1,2,3,4] Reecha Sofat,[4,6] Valerie Kuan,[1,2,3] Claudia Langenberg,[7,8,9] Harry Hemingway,[2,3,4,10] Joan K Morris,[11] Nicholas J Wald[10,11]

ABSTRACT

OBJECTIVE To clarify the performance of polygenic risk scores in population screening, individual risk prediction, and population risk stratification.

DESIGN Secondary analysis of data in the Polygenic Score Catalog.

SETTING Polygenic Score Catalog, April 2022. Secondary analysis of 3915 performance metric estimates for 926 polygenic risk scores for 310 diseases to generate estimates of performance in population screening, individual risk, and population risk stratification.

PARTICIPANTS Individuals contributing to the published studies in the Polygenic Score Catalog.

MAIN OUTCOME MEASURES Detection rate for a 5% false positive rate (DR_5) and the population odds of becoming affected given a positive result; individual odds of becoming affected for a person with a particular polygenic score; and odds of becoming affected for groups of individuals in different portions of a polygenic risk score distribution. Coronary artery disease and breast cancer were used as illustrative examples.

RESULTS For performance in population screening, median DR_5 for all polygenic risk scores and all diseases studied was 11% (interquartile range 8-18%). Median DR_5 was 12% (9-19%) for polygenic risk scores for coronary artery disease and 10% (9-12%) for breast cancer. The population odds of becoming affected given a positive results were 1:8 for coronary artery disease and 1:21 for breast cancer, with background 10 year odds of 1:19 and

1:41, respectively, which are typical for these diseases at age 50. For individual risk prediction, the corresponding 10 year odds of becoming affected for individuals aged 50 with a polygenic risk score at the 2.5th, 25th, 75th, and 97.5th centiles were 1:54, 1:29, 1:15, and 1:8 for coronary artery disease and 1:91, 1:56, 1:34, and 1:21 for breast cancer. In terms of population risk stratification, at age 50, the risk of coronary artery disease was divided into five groups, with 10 year odds of 1:41 and 1:11 for the lowest and highest quintile groups, respectively. The 10 year odds was 1:7 for the upper 2.5% of the polygenic risk score distribution for coronary artery disease, a group that contributed 7% of cases. The corresponding estimates for breast cancer were 1:72 and 1:26 for the lowest and highest quintile groups, and 1:19 for the upper 2.5% of the distribution, which contributed 6% of cases.

CONCLUSION Polygenic risk scores performed poorly in population screening, individual risk prediction, and population risk stratification. Strong claims about the effect of polygenic risk scores on healthcare seem to be disproportionate to their performance.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Disagreement exists on the performance of polygenic risk scores in screening, prediction, and risk stratification, and therefore their potential value in healthcare

WHAT THIS STUDY ADDS

⇒ Calculating informative performance metrics for 926 polygenic risk scores for 310 diseases from the Polygenic Score Catalog indicated poor performance in screening (median detection rate for a 5% false positive rate of 11%), with correspondingly poor performance in individual risk prediction and population risk stratification

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

⇒ The wide scope and analytical approach of this study should resolve disagreement on the value of polygenic risk scores and avoid unjustified expectations about their role in preventing disease

Introduction

A polygenic risk score represents the weighted sum of independent DNA sequence variants present in an individual's genome that are associated with the risk of a particular disease.[1] The weight assigned to each variant is based on the strength of its disease association, estimated from a genome wide association study. The increasing range and scale of genome wide association studies over the past decade, now spanning more than 2500 diseases or traits,[2] has proliferated polygenic risk scores, with widespread interest in potential healthcare applications[3] and attention from policy makers.[4]

Claims have been made that polygenic risk scores generate "substantial" improvements in risk prediction,[5] will "power a transformative change to healthcare",[6] and should be made ready to implement in practice.[7] [8] In a move towards clinical implementation, position papers have been published on reporting standards and responsible clinical use from the Clinical Genome Resource (ClinGen) Complex Disease Working Group[9] and the Polygenic Risk Score Task Force of the International Common

Disease Alliance.[10] Individual consumers and healthcare providers can already access commercial genetic testing and software services based on polygenic scores.[11–13] A "world first" pilot trial of predictive genetic testing for cardiovascular disease is also underway in participants attending vascular health checks in the NHS,[14] and genetic risk scores for prediction of disease are central to the aims of the Our Future Health programme aiming to recruit five million UK adults.[15]

These claims, however, are disputed[16–19] resulting in disagreement on the performance of polygenic risk scores in population screening, individual risk prediction, and population risk stratification, making their role in medicine and public health uncertain. Recently, Lambert and colleagues produced the Polygenic Score Catalog, a comprehensive, regularly updated, open access directory of studies on polygenic scores for quantitative traits (eg, blood pressure) and polygenic risk scores for diseases (eg, breast cancer).[20 21] The catalogue lists the performance metrics for polygenic risk scores as hazard ratios or odds ratios for an increment of one standard deviation in the score, area under the receiver operating characteristic curve, or the C index.

But these metrics are not directly informative of performance in population screening, individual risk prediction, and population risk stratification. Screening is defined as "the systematic application of a test or enquiry to identify individuals at sufficient risk of a specific disorder to benefit from further investigation or direct preventive action, among persons who have not sought medical attention because of symptoms of that disorder."[22] Individual risk prediction is the estimation for an individual of the risk of becoming affected in a given time frame, and population risk stratification is the estimation for a population subgroup of the absolute risk (or odds) of becoming affected in a given time frame.

An appropriate performance measure is the odds of becoming affected, which is the positive predictive value expressed as an odds. For example, an odds of 1:9 equates to a risk of one in 10 or 10%. The odds of becoming affected is calculated by multiplying the background odds of developing disease in a specified time frame by the likelihood ratio associated with a positive test (population screening), with a particular polygenic score value (individual risk prediction), or occupancy of a particular polygenic risk score quantile group (population risk stratification).

We mathematically derived the odds of becoming affected in each of these scenarios based on the metrics reported in the Polygenic Score Catalog. We used breast cancer and coronary artery disease as illustrative examples and scrutinised two proposed early clinical uses of polygenic risk scores: to improve on the performance of the established risk factor models in the prediction of coronary artery disease and stroke, and to prioritise mammographic

screening at a younger age for the detection of breast cancer.

Methods

Reported performance metrics

By April 2022, the Polygenic Score Catalog had curated 13 828 performance metric estimates for 2194 polygenic scores (unique polygenic score codes), for 544 diseases or traits (unique experimental factor ontology identifiers), reported in 303 unique publications. We removed polygenic scores for continuous traits and those with implausible values (167 instances where the hazard ratio or odds ratio for one standard deviation was recorded as <1, two instances where the area under the receiver operating characteristic curve was <0.5, and one instance where the C index was recorded as 632), leaving 3915 performance metric estimates for 926 polygenic risk scores with 310 unique binary outcomes (mainly diseases). The reported performance metrics were odds ratio for one standard deviation in 1216 instances, hazard ratio for one standard deviation in 378, area under the receiver operating characteristic curve in 2077, and C index in 244 instances (online supplemental file 1).

Converting reported metrics to useful performance metrics

Polygenic risk scores have a gaussian distribution with the same standard deviation in affected and unaffected groups (online supplemental file 2). We used these properties together with the metrics reported in the Polygenic Score Catalog to mathematically derive measures that are more useful for judging the performance of polygenic risk scores in their intended applications.

The first step in this calculation is to use the reported metrics to calculate the difference in mean values for polygenic risk scores between affected and unaffected groups. The calculations have been described previously[23–25] and are explained in detail in online supplemental file 2. We used the calculated difference in mean values to determine the overlap in the distributions of the polygenic risk score between the affected and unaffected groups. This method allows calculation of the detection rate and false positive rate, which are the percentage of people with a polygenic score above a particular cut-off value (positive test) among those who are later affected or remain unaffected by disease, respectively. For simplicity and consistency, we set the polygenic score cut-off value at the 95th centile for the unaffected group (1.645 standard deviation units from the mean). This cut-off value defines a 5% false positive rate, and the corresponding detection rate (DR_5) is the detection rate for a 5% false positive rate.[23 25] The work sheets in online supplemental file 1 can be used to enter the odds ratio or hazard ratio for one

standard deviation, area under the receiver operating characteristic curve, or the C index and return the DR_5 value. A risk screening converter that returns the detection rate for any user defined false positive rate value is available at https://www.medicalscreeningsociety.com/rsc.asp.

The likelihood ratio in screening is the ratio of the detection rate/false positive rate. In individual risk prediction, the likelihood ratio is the ratio of the heights of the gaussian distribution curves for affected and unaffected individuals at a particular polygenic risk score centile. In risk stratification, the likelihood ratio is the ratio of areas under the relative frequency distributions for affected and unaffected individuals in each polygenic score quantile (eg, each fifth of the polygenic score distribution; figure 1 and online supplemental file 2). In each case, multiplying the likelihood ratio by the background odds of disease for the whole population gives the corresponding odds of becoming affected for the individual or group of interest. When we discuss a polygenic score centile or quantile, we are referring to the distribution in the unaffected group. When referring to a particular polygenic risk score, we used the Polygenic Score Catalog identifier number.

We reanalysed tabular data taken from two original sources[5][26] to quantify the extent to which the addition of information on polygenic risk score to data on conventional risk factors improved the prediction of coronary artery disease and stroke. For this analysis, we extracted the reported counts of affected and unaffected individuals above and below the 10 year risk cut-off values recommended in guidelines for starting treatment with statins. We did this analysis separately for the counts reported by the authors for conventional risk factor models, and for the conventional risk factor models with the addition of polygenic risk scores (online supplemental file 1). We used these counts to calculate the detection rate and false positive rate with and without information on the polygenic risk score. We then calculated the number of individuals who need to be genotyped (and a polygenic score calculated) to detect or prevent one additional coronary artery disease event or stroke with the authors own assumptions about the risk reduction from treatment with statins. We refer to this value as the number needed to genotype (table 1). We modelled the use of a breast cancer polygenic risk score to prioritise mammographic screening at age 40 rather than from the currently recommended age of 50.

Patient and public involvement
Patients and the public were not involved in the design, or conduct, or reporting of this research. We plan to work with patients and the public to disseminate the findings through the patient and public representative groups of the Multimorbidity Mechanisms and Therapeutics Research Collaborative and the UCL Hospitals NIHR Biomedical Research Centre.

Results
Performance of polygenic risk scores in screening
For all diseases studied, median DR_5 based on all polygenic risk scores was 11% (interquartile range 8-18%); that is, 89% (82-92%) of patients were missed. Median DR_5 values for polygenic risk scores whose performance was reported with odds ratio or hazard ratio for one standard deviation were 9% (6-12%) and 8% (7-10%), respectively. For polygenic risk score performance reported based on area under the receiver operating characteristic curve or the C index, median DR_5 values were 14% (10-22%) and 19% (13-25%), respectively. Figure 2 shows median DR_5 values for polygenic risk scores for 28 common diseases, including coronary artery disease and breast cancer.

Coronary artery disease
Median DR_5 from performance metrics for 27 polygenic risk scores for coronary artery disease was 12% (interquartile range 9-20%; 88% of patients missed) (figure 3), corresponding to a likelihood ratio of 2.4. Applied at age 50, with a background 10 year risk of coronary artery disease of 5% (odds 1:19), the odds of becoming affected given a positive result were 2.4:19 or about 1:8; that is, false positive results outnumbered true positive results by about eight to one. Reducing the cut-off value to reduce the false positive rate to 1% reduced the detection rate to 3%, with 97% of patients missed. Retaining a 5% false positive rate but applying the test in a population with a risk of coronary artery disease of one in 56 (about 2%), over the same period (background odds 1:55; eg, at about age 40 years of age), gave an odds of becoming affected given a positive result of 1:23, with false positive results outnumbering true positive results by just over 20 to one.

Breast cancer
Median DR_5 from performance metrics reported for 108 polygenic risk scores for breast cancer was 10% (interquartile range 9-12%; 90% of patients missed) (figure 3), corresponding to a likelihood ratio of 2. Applied at age 50, with a background 10 year risk of breast cancer of about 2.5% (odds 1:41), the odds of becoming affected given a positive result were 1:21. Applying the polygenic risk score as a test at age 40, when the background 10 year odds was 1:64, gave an odds of becoming affected given a positive result of 1:32, with false positive results outnumbering true positive results by just over 30 to one.
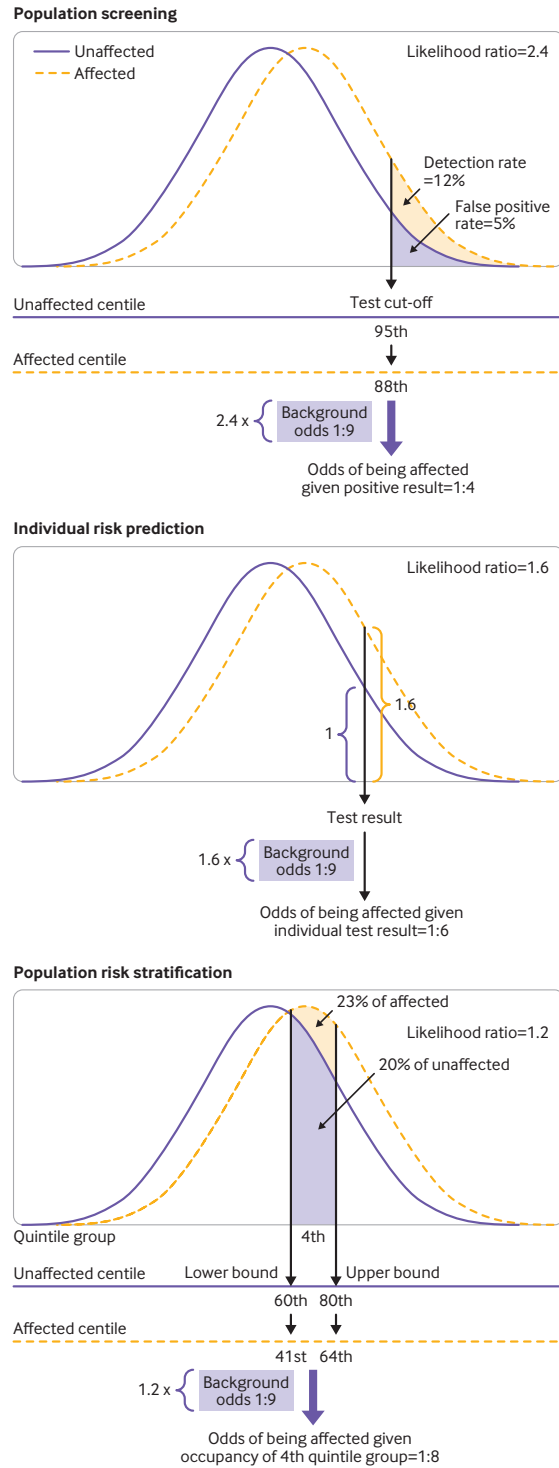
Figure 1 | Derivation of metrics useful in assessing performance of polygenic risk scores in population screening, individual risk prediction, and population risk stratification. Difference in mean values for polygenic risk scores between affected and unaffected groups (and standard deviations) allows determination of overlap in polygenic risk score distributions between the two groups. Likelihood ratio in screening is detection rate for a specified false positive rate (5%) and is the ratio of the shaded areas in the top panel. In individual risk prediction, likelihood ratio is the ratio of the heights of the distributions at a specified polygenic risk score (middle panel). In population risk stratification, likelihood ratio is the ratio of areas under the distributions for a specified group of the population (fourth quintile group in bottom panel). Multiplication of the likelihood ratio by the background odds of disease in the population (1:9) allows calculation of the odds of becoming affected for each patient

Table 1 | Effect of adding polygenic risk score to non-genetic risk factors in prediction of coronary artery disease and stroke

Study	Population screened (cardiovascular events)	Risk assessment	Detection rate (%)	False positive rate (%)	No of patients prescribed statin	Cardiovascular events above cut-off value	Events prevented	Number needed to genotype
Sun et al[26]	100 000 (7997)	Conventional risk factors	60	24	26 722	4783	957	—
		Conventional risk factors+polygenic risk score	61	23	26 445	4870	974	5882
Riveros et al[5]	186 451 (4247)	QRISK3	81	42	79 754	3450	690	—
		QRISK3+polygenic risk score	84	41	78 092	3557	711	8879

Values are based on data from Sun et al[26] and Riveros-Mckay et al.[5] Both studies used data from UK Biobank. Sun et al developed a conventional risk factor model and examined the effect of adding polygenic risk scores for coronary artery disease (Polygenic Score Catalog identifier PGS000018) and stroke (PGS000039) on the prediction of subsequent coronary artery disease and stroke events. Sun et al used a 10 year risk cut-off value of 10% for prescribing treatment with statins. Riveros-McKay et al[5] modelled screening performance in 186 451 participants based on the cardiovascular risk score, QRISK3, also with a 10% risk cut-off value for prescription of statins. Data on events reported by Riveros-Mackay et al[5] were for coronary artery disease only rather than coronary artery disease and stroke. Calculations assume that all those exceeding the specified risk cut-off value receive a statin, 100% adherence, and that statin treatment produces a 20% reduction in relative risk. Counts were used to calculate number needed to genotype: number of individuals that need to be genotyped (and have a polygenic risk score calculated) to detect or prevent one additional cardiovascular event (see text).

Performance of polygenic risk scores in individual risk prediction

The overlap in distributions of polygenic scores derived from the metrics in the Polygenic Score Catalog allowed calculation of the likelihood ratio for an individual which, together with the background odds of the disorder for the population, can be used to calculate the odds of becoming affected for that individual (online supplemental file 2).[25]

Coronary artery disease

The odds of developing coronary artery disease in the next 10 years were 1:54, 1: 29, 1:15, and 1:8 with a polygenic risk score at the 2.5th, 25th, 75th, and 97.5th centiles, respectively, at age 50 (when the background odds was 1:19) (figure 4), and 1:157, 1:85, 1:45, and 1:24, respectively, at age 40 (when the background odds was 1:55).

Breast cancer

The average 10 year odds of breast cancer was 1:41 for a woman aged 50 and 1:64 for a woman aged 40. The corresponding odds of being affected were 1:91, 1:56, 1:34, and 1:21, respectively, at age 50 (figure 4), and 1:142, 1:88, 1:53, and 1:33 at age 40, for a woman with a polygenic risk score at the 2.5th, 25th, 75th, and 97.5th centiles, respectively.

Performance of polygenic risk scores in risk stratification

Coronary artery disease

Figure 5 shows the overlapping distributions for affected and unaffected individuals applied to a hypothetical cohort of 100 000 men aged 50 years grouped into polygenic risk score quintile groups. The 10 year odds of coronary artery disease were reduced from the average of 1:19 for all men to 1:41
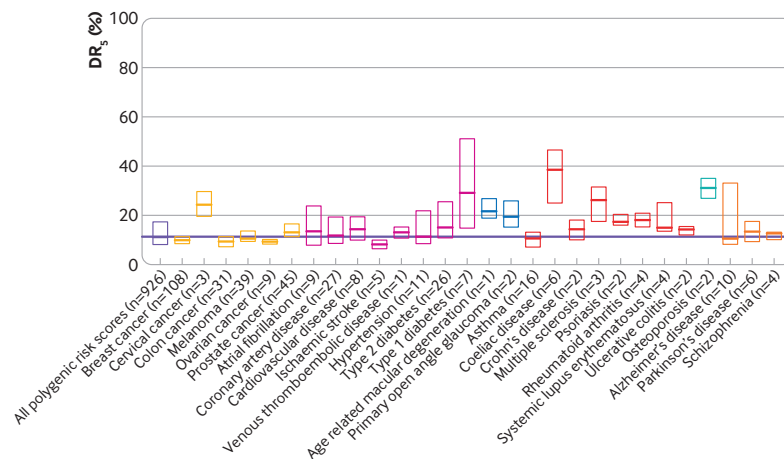

Figure 2 | Performance in screening estimated for polygenic risk scores included in the Polygenic Score Catalog from April 2022. Limits of each box represent interquartile range and horizontal line within each box is estimated detection rate for a 5% false positive rate (DR_5) based on performance metrics reported for corresponding polygenic risk scores. Selected diseases are colour coded into categories cancers, cardiometabolic conditions, ocular diseases, allergic or autoimmune diseases, bone disease, and neuropsychiatric diseases. Horizontal line is estimated median DR_5 value based on performance metrics for all 926 polygenic risk scores and all diseases studied in the Polygenic Score Catalog

Coronary artery disease
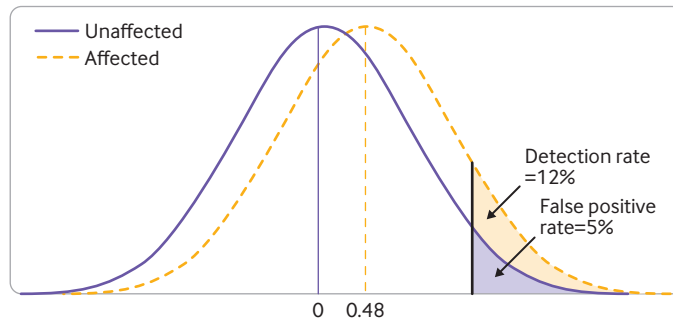
Odds ratio top *v* bottom 1%	13
Odds ratio top *v* bottom 5%	7
Odds ratio top *v* bottom 10%	5
Odds ratio top *v* bottom 20%	4
Odds ratio top *v* bottom 25%	3
Detection rate (%)	12
False positive rate (%)	5

Breast cancer
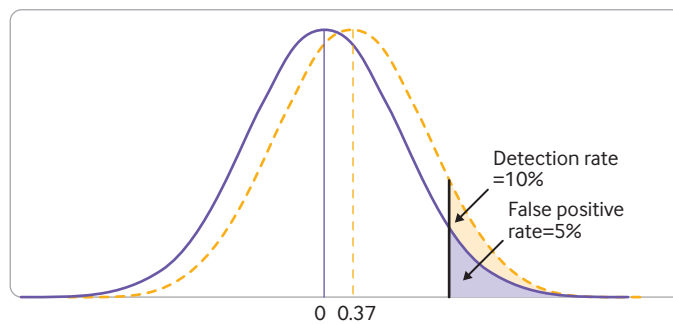
Odds ratio top *v* bottom 1%	7
Odds ratio top *v* bottom 5%	5
Odds ratio top *v* bottom 10%	4
Odds ratio top *v* bottom 20%	3
Odds ratio top *v* bottom 25%	3
Detection rate (%)	10
False positive rate (%)	5

Figure 3 | Relative polygenic risk score distributions among those later affected or not by coronary artery disease and breast cancer. Mean value of polygenic risk score distribution in those later affected was shifted 0.48 standard deviation units to the right of the mean of the distribution for those who remained unaffected by coronary artery disease, and 0.37 standard deviation units to the right for breast cancer. Also shown are corresponding values for detection rate for a 5% false positive rate (DR_5) and for odds ratios (rounded to the nearest whole number) for comparisons of top and bottom 1%, 5%, 10%, 20%, and 25% of unaffected polygenic risk score distribution

for those in the lowest quintile group, and increased to 1:11 in the highest quintile group. Figure 6 focuses on those at the highest risk, and the 10 year odds were 1:7 for the upper 2.5% of the polygenic risk score distribution for coronary artery disease, but this group contributed only 7% of patients.

Breast cancer
We applied the same approach to a hypothetical cohort of women aged 50 with a background 10 year odds of breast cancer of 1:41 (figure 5). The odds reduced by about half to 1:72 for those in the lowest

quintile group and almost doubled to about 1:26 for those in the highest quintile group of the polygenic risk score distribution. For the highest 2.5% of the polygenic risk score distribution (figure 6), the odds were increased to 1:19, but the latter group only accounted for 6% of patients with breast cancer.

Polygenic risk scores for screening along with conventional risk factors or tests
Coronary artery disease
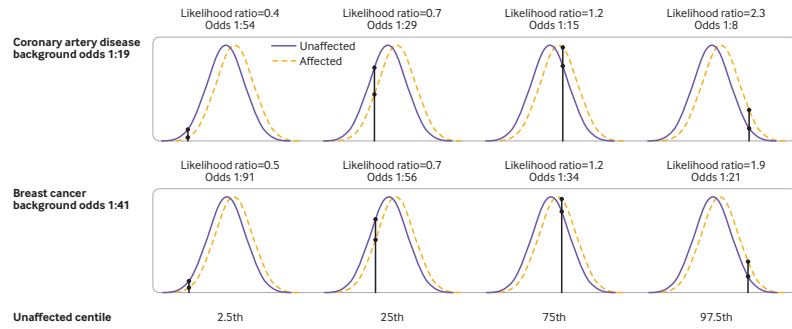Adding polygenic risk scores to conventional risk factors (eg, blood pressure and low density

Figure 4 | Likelihood ratios and 10 year odds of coronary artery disease and breast cancer for people aged 50 with a polygenic risk score result corresponding to 2.5th, 25th, 75th, and 97.5th centiles of the corresponding distribution

lipoprotein cholesterol) has been proposed to usefully improve coronary artery disease and stroke screening to indicate who should be prescribed a statin for primary prevention. Table 1 shows the results from Sun et al,[26] applied in a hypothetical cohort of 100 000 individuals aged 40 with a risk factor profile representative of the English population and a background 10 year risk of coronary artery disease and stroke of 8%. A conventional, multi-risk factor model incorporating age and a 10 year risk

cut-off value of 10%, detected 60% of those later affected by coronary artery disease or stroke at a false positive rate of 24% (DR_{24}=60%). The addition of polygenic risk scores for coronary artery disease and stroke to the model (Polygenic Score Catalog identifiers PGS000018 and PGS000039, respectively) detected 61% of those affected with a false positive rate of 23% (DR_{23}=61%). Assuming a 10 year risk cut-off value of 10% for prescribing statins,[27] 100% adherence, and adopting the assumption of Sun et al
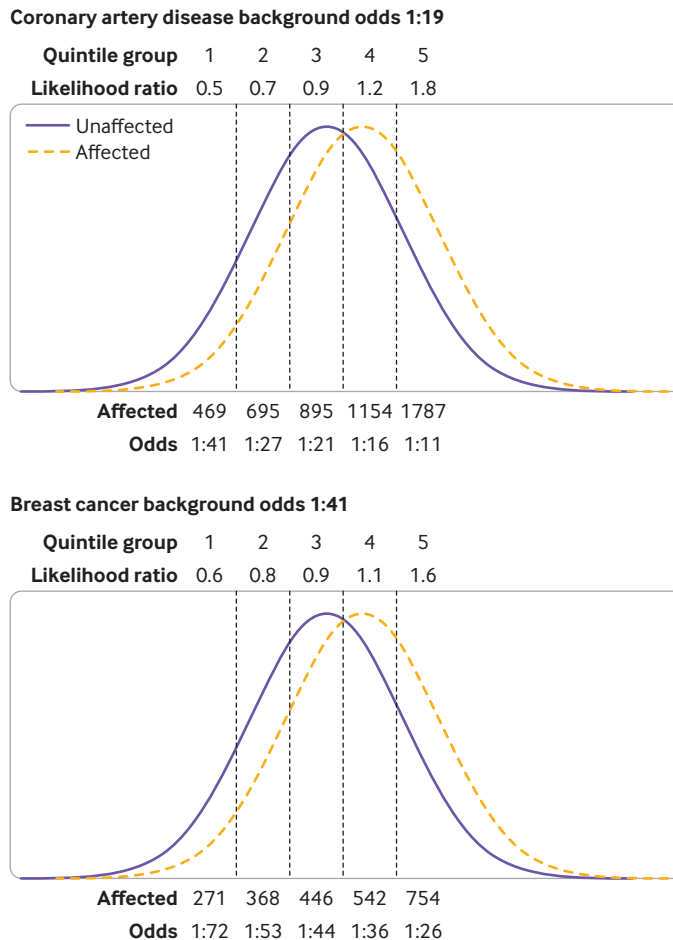

Figure 5 | Likelihood ratios, odds, and number of affected and unaffected individuals for each quintile group in a hypothetical population of 100 000 individuals with a background 10 year odds of coronary artery disease of 1:19, and women with a 10 year odds of breast cancer of 1:41

Coronary artery disease background odds 1:19

Group	Lowest 2.5%	Middle 95%	Highest 2.5%
Likelihood ratio	0.30	0.97	2.76

Affected	37	4618	345
Unaffected	2375	90 250	2375
Odds	1:64	1:20	1:7
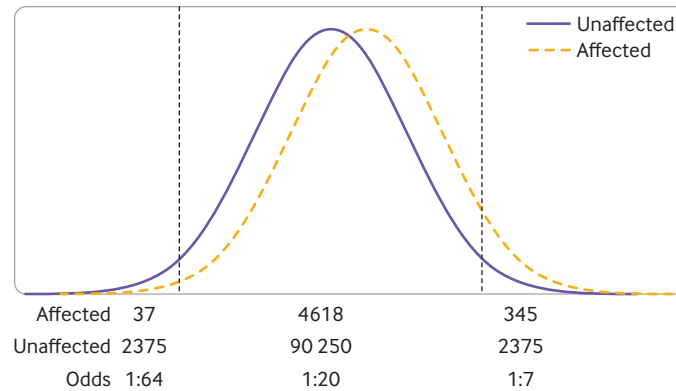
Breast cancer background odds 1:41

Group	Lowest 2.5%	Middle 95%	Highest 2.5%
Likelihood ratio	0.40	0.98	2.21

Affected	24	2225	132
Unaffected	2440	92 738	2440
Odds	1:102	1:42	1:19
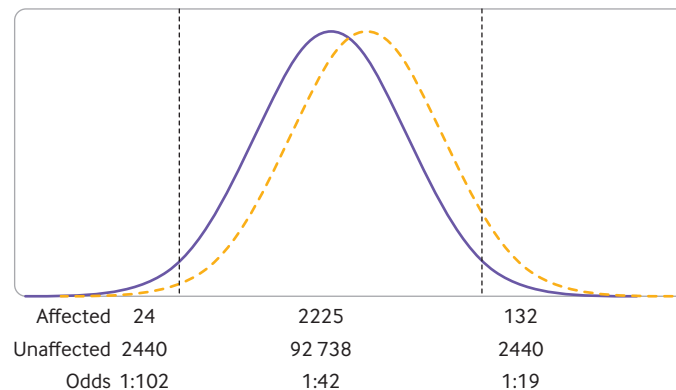
Figure 6 | Likelihood ratios and 10 year odds of coronary artery disease and breast cancer for people aged 50 comparing highest and lowest 2.5% of the unaffected polygenic risk score distributions

that statins reduce the risk of coronary artery disease and stroke by 20%,[26] 974 events would be prevented with a model based on conventional risk factors and polygenic risk scores compared with 957 with a conventional risk factor model and no genetic information, a gain of 17 patients prevented (table 1).

This method gives a number needed to genotype to prevent one additional event of 5882. Sun et al also estimated that 1029 coronary artery disease and stroke events would be prevented with a hybrid model, when conventional assessment of risk factors is followed by polygenic risk scores only for those with an intermediate (5-10%) 10 year risk. Replacing this more complicated model, however, with one where the whole cohort receives statins would prevent 1600 cardiovascular events based on the same assumptions (online supplemental file 1). Because age is a major determinant of the risk of coronary artery disease and stroke, age alone performs about as well as multiple risk factor models that include age.[28] Based on the rarity of coronary artery disease and

stroke events at age <50, an age cut-off of 50 instead of 40 would prevent almost as many events but with fewer false positive results.[29]

Similar results were obtained with Riveros-McKay et al's data.[5] These authors also investigated the extent to which the addition of a polygenic risk score to conventional risk factors improved the identification of UK Biobank participants eligible to receive statins because their 10 year risk of coronary artery disease and stroke exceeded the cut-off values used in UK or US primary prevention guidelines. Deriving the appropriate metrics from their data (table 1 and online supplemental file 1) clarifies the effect of adding information from a polygenic risk score for coronary artery disease. With a 10 year risk cut-off value of 10% for starting statins, the cardiovascular risk score, QRISK3 model, based on conventional risk factors including age, detected 81% of patients at a false positive rate of 42% (DR_{42}=81%). The addition of a polygenic risk score to the model detected 84% of patients for a false positive rate of 41% (DR_{41}=84%).
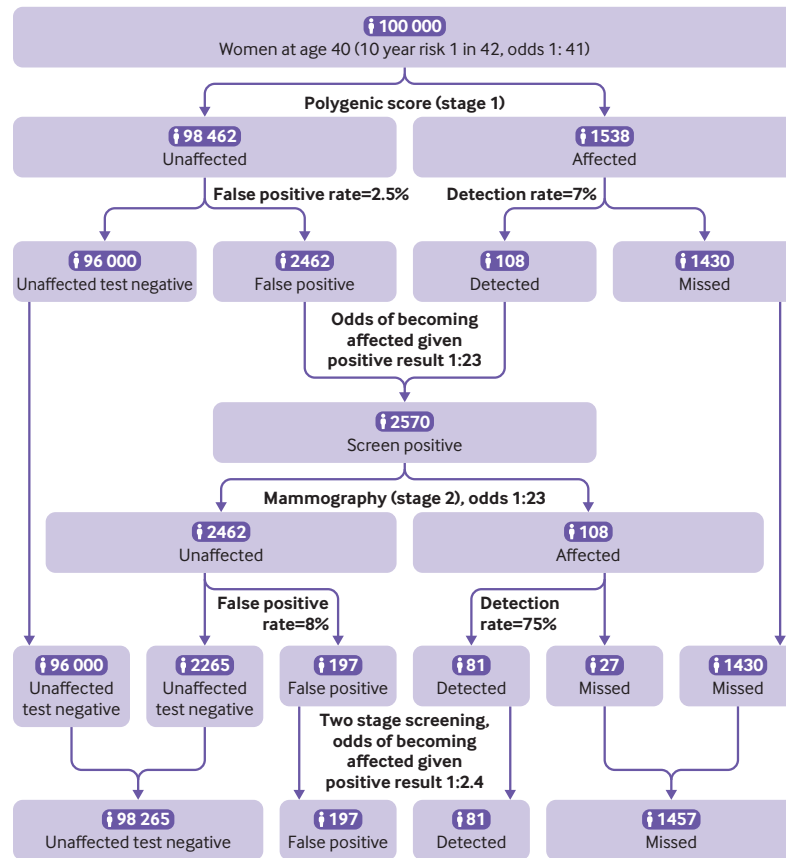
Figure 7 | Estimated number of patients with breast cancer detected and missed, number of false positive results, and number of additional mammograms for a two stage screening test with a polygenic risk score (Polygenic Score Catalog identifier PGS000004) with a cut-off value at the unaffected 97.5th centile. Estimates are based on a hypothetical cohort of 100 000 women aged 40 with a background 10 year odds of breast cancer of 1:41. Performance of mammography in the detection of breast cancer uses estimates from the literature[31]

Based on the authors' assumption that statins reduce coronary artery disease and stroke events by 20%, 711 events would be prevented with a model based on conventional risk factors and polygenic risk scores compared with 690 with a conventional risk factor model and no genetic information, a gain of 21 patients prevented. This calculation gives a number needed to genotype to prevent one additional event based on this study of 8879.

Breast cancer

Using polygenic risk scores to prioritise the use of established screening tests for cancer has also been proposed.[3] One suggestion is that younger women should undergo mammographic screening if their risk of breast cancer, determined with a polygenic risk score, exceeds that of an average woman aged 50, the age when mammography is offered to all women. Figure 6 shows that women aged 40 at or above the unaffected 97.5th centile of a breast cancer polygenic risk score distribution have an odds of breast cancer of 1:19, higher than the average 10 year odds at age 50 of 1:41.[30] Figure 7 shows that by using the breast cancer polygenic risk score (Polygenic Score Catalog identifier PGS000004) as a stage 1 screen

in 100 000 women aged 40, applying the unaffected 97.5th centile as a cut-off value would result in 2570 women with a high risk polygenic score being offered mammography, of whom 108 would be affected and 2462 unaffected (odds of becoming affected given a positive result 1:23). Assuming 100% uptake and a DR_8 value of 75%,[31] mammography would then correctly identify 81 of the 108 affected individuals but miss 27 patients with breast cancer. However, 1430 patients with breast cancer (over 10 times as many) are estimated among the 97 430 women aged 40 with a polygenic risk score below the unaffected 97.5th centile who would not be offered mammography.

Discussion
Principal findings
Our results showed the poor performance of polygenic risk scores in population screening, individual disease prediction, and population risk stratification. This finding is not obvious from the metrics reported in the Polygenic Score Catalog but is clear based on the appropriate metrics used in this study. Our conclusion is consistent with that of others,[16 17 19] but is insufficiently recognised. The findings are relevant

to consumers, patients, doctors, those involved in preventive medicine and public health, as well as funders and policy makers.

Polygenic risk score distributions overlapped substantially for all conditions studied, and this extensive overlap constrained their performance in each of their intended applications, whether used alone or in combination with conventional risk factors or screening tests. For example, achieving a clinically useful performance in population screening, such as an 80% detection rate for a 5% false positive rate (DR_5=80%) requires an odds ratio for one standard deviation of 12 or higher (compared with the median observed value of 1.31) or an area under the receiver operating characteristic curve of 0.96 (compared with the median observed value of 0.65). Only 11.4% of the area under the curve values in the Polygenic Score Catalog exceeded 0.8, which equates to a DR_5 of 32%, with most of these resulting from large effect variants at the HLA locus in a few autoimmune diseases (figure 1 and online supplemental file 1).

Study implications

When a risk factor has a monotonic relation with risk of disease,[32] more instances arise among the majority with near average risk factor values than among the few with more extreme values, termed the prevention paradox.[33 34] In this respect, polygenic risk scores are similar to some non-genetic risk factors, such as blood pressure and low density lipoprotein cholesterol, which although causal, are poor predictors of coronary artery disease.[16 35] That the performance of polygenic risk scores in the prediction of coronary artery disease is sometimes compared favourably with that of blood pressure and cholesterol[26] is to benchmark one poor predictor against another.

Where safe and inexpensive preventive interventions are available (eg, statins and blood pressure lowering drugs for prevention of coronary artery disease and stroke), broadening rather than limiting eligibility for such interventions gives greater public health benefits.[36] Prevention of coronary artery disease and stroke has been achieved in effect by the progressive lowering of the 10 year risk cut-off value for prescription of statins in primary prevention. The cut-off value was reduced from a 10 year risk of coronary artery disease in the UK in 1997 of 30%,[37] to 10% for the 10 year risk of coronary artery disease or stroke in the UK from 2016[27] and to 7.5% in the US from 2019.[38] The reduction in the risk cut-off value resulted from reduced drug acquisition costs through patent expiry, and by accumulating evidence on long term safety. Eligibility could be extended even further and simplified by using age alone to guide prescription of statins for primary prevention, preventing coronary artery disease and stroke in many more patients.[28] In contrast, retaining the same 10 year risk cut-off value and adding information on polygenic risk score to conventional risk factor

models has a much weaker effect. Based on recently reported data,[5 26] we showed that several thousand individuals need to be genotyped and a polygenic risk score calculated to prevent one additional vascular event.

Identifying a minority of individuals at very high risk (with genetics or other means) might be justified if a preventive intervention is costly, resource limited, or has substantial harms.[39] With breast cancer as an example, however, we showed that identifying those at high risk requires testing in all and, apart from missing the many more patients among those at average risk, generates many false positive results. This finding could have substantial downstream resource implications for healthcare systems if, for example, genetic risk stratification was followed by a confirmatory screening test, such as mammography for breast cancer.[40] In this case, reducing the age cut-off value for mammography for all women without determining their polygenic risk score might be more sensible.

The enthusiasm surrounding polygenic risk scores might have been encouraged by pressure on academia to demonstrate a tangible health effect after decades of research investment in human genomics and by commercial opportunity. Unrealistic expectations have probably been raised by use of uninformative metrics. Publications on polygenic risk scores often illustrate comparisons between mutually exclusive groups (eg, those in opposite ends of a polygenic score distribution).[41] This finding is relevant in aetiological studies but is not relevant in screening. Figure 3 shows seemingly impressive odds ratios of 13, 7, 5, 4, and 3 for comparisons of the top versus the bottom 1%, 5%, 10%, 20%, and 25%, respectively, of the polygenic risk score distribution for coronary artery disease, all reduced to a DR_5 of only 12%. What is relevant in screening is the risk of an event in a group compared with that of the whole population, which is achieved with the calculation of the detection rate for a specified false positive rate.

Policy implications

Our findings are relevant to commercial providers of genetic tests and to researchers working on polygenic risk scores. Commercial providers could communicate individual test results to customers with greater clarity and relevance to performance in disease prediction; for example, by presenting the overlapping distributions of polygenic risk scores among those later affected and unaffected and by presenting an absolute measure of risk for an individual or group, which requires additional information on population average risk at a particular age over a specified time. At the same time, as already suggested,[42] policy makers might wish to consider stricter regulation of commercial genetic tests based on polygenic risk scores, with a focus on clinical performance and not just assay performance (as indicated by the Royal Statistical Society Diagnostic Tests Working Group Report[43]), to protect the public from

unrealistic expectations and already stretched public health systems from becoming overburdened by the management of false positive results. Researchers reporting studies on polygenic risk scores should present as a minimum: mean and standard deviation values for polygenic risk scores among later affected and unaffected individuals; overlap in their distributions; relevant performance metrics, such as the detection rate for a specified false positive rate (eg, DR_5), avoiding the need to make this calculation indirectly[23]; and performance of polygenic risk scores with and without the inclusion of other variables so that users can judge the incremental benefit provided by the polygenic risk score itself.

Although our analysis showed the poor performance of polygenic risk scores in screening, prediction, and risk stratification, these scores might be useful in other situations. For example, polygenic scores might explain the variable penetrance of rare mutations in monogenic diseases (eg, hypertrophic cardiomyopathy or familial hypercholesterolaemia), and be used to help detect patients. Other predictive applications of genotyping also exist, for example in pharmacogenetic testing to optimise the efficacy and safety of medicines. Genotyping might also be of value in blood and tissue matching. Because genetic variation is transmitted from parents to offspring through a randomised process (like treatment allocation in a clinical trial), and is unaltered by disease, an important translational application arising from genomic discoveries could be providing evidence on disease causation and targets for pharmaceutical intervention.[44]

Conclusion

Use of the appropriate metrics showed poor performance of polygenic risk scores in population screening, individual risk prediction, and population risk stratification. The wide scope and analytical approach of our study might help to resolve the debate on the value of polygenic risk scores, and avoid unjustified expectations about their role in the prediction and prevention of disease.

AUTHOR AFFILIATIONS

[1]Institute of Cardiovascular Science, University College London, London, UK
[2]British Heart Foundation Research Accelerator, University College London, London, UK
[3]National Institute of Health Research Biomedical Research Centre, University College London Hospitals, London, UK
[4]Health Data Research UK, London, UK
[5]University Medical Centre Utrecht, Utrecht, Netherlands
[6]Department of Pharmacology and Therapeutics, University of Liverpool, Liverpool, UK
[7]Precision Healthcare University Research Institute, Queen Mary University of London, London, UK
[8]Computational Medicine, Berlin Institute of Health at Charite Universitatzmedizin, Berlin, Germany
[9]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
[10]Institute of Health Informatics, University College London, London, UK
[11]Population Health Research Institute, St George's University of London, London, UK

ORCID iD
Jasmine Gratton http://orcid.org/0000-0002-9781-9762

REFERENCES

1 Wray NR, Lin T, Austin J, et al. From basic science to clinical application of polygenic risk scores: a primer. JAMA Psychiatry 2021;78:101–9. 10.1001/jamapsychiatry.2020.3049

2 Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol 2019;2:9. 10.1038/s42003-018-0261-x

3 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet 2018;19:581–90. 10.1038/s41576-018-0018-x

4 GOV.UK. Genome UK: the future of healthcare. Available: https://www.gov.uk/government/publications/genome-uk-the-future-of-healthcare [Accessed 24 Jan 2022].

5 Riveros-Mckay F, Weale ME, Moore R, et al. Integrated polygenic tool substantially enhances coronary artery disease prediction. Circ Genom Precis Med 2021;14:e003304. 10.1161/CIRCGEN.120.003304

6 For health systems - Genomics plc, Available: https://www.genomicsplc.com/precision-health/health-systems/ [Accessed 25 Aug 2022].

7 Knowles JW, Ashley EA. Cardiovascular disease: the rise of the genetic risk score. PLoS Med 2018;15:e1002546. 10.1371/journal.pmed.1002546

8 Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet 2019;28:R133–42. 10.1093/hmg/ddz187

9 Wand H, Lambert SA, Tamburro C, et al. Improving reporting standards for polygenic scores in risk prediction studies. Nature 2021;591:211–9. 10.1038/s41586-021-03243-6

10 Polygenic Risk Score Task Force of the International Common Disease Alliance. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nat Med 2021;27:1876–84. 10.1038/s41591-021-01549-6

11 Horton R, Crawford G, Freeman L, et al. Direct-to-consumer genetic testing. BMJ 2019;367:l5688. 10.1136/bmj.l5688

12 Genomics PLC. Available: https://www.genomicsplc.com/ [Accessed 5 Jun 2022].

13 Allelica | Polygenic risk score. Available: https://www.allelica.com/ [Accessed 5 Jun 2022].

14 The Guardian. NHS pilots genetic testing to predict risk of heart disease | Heart disease, Available: https://www.theguardian.com/society/2022/jan/17/nhs-pilots-genetic-testing-to-predict-risk-of-heart-disease [Accessed 11 Sep 2022].

15 BBC News. Health of nation study calls on millions to sign up. Available: https://www.bbc.co.uk/news/health-63331805 [Accessed 25 Oct 2022].

16 Wald NJ, Old R. The illusion of polygenic disease risk prediction. Genet Med 2019;21:1705–7. 10.1038/s41436-018-0418-5

17 Groenendyk JW, Greenland P, Khan SS. Incremental value of polygenic risk scores in primary prevention of coronary heart disease: a review. JAMA Intern Med 2022;182:1082. 10.1001/jamainternmed.2022.3171

18 Sud A, Turnbull C, Houlston R. Will polygenic risk scores for cancer ever be clinically useful? NPJ Precis Oncol 2021;5:40. 10.1038/s41698-021-00176-1

19 Sud A, Horton RH, Hingorani AD, et al. Realistic expectations are key to realising the benefits of polygenic scores. BMJ 2023;380:e073149. 10.1136/bmj-2022-073149

20 Lambert SA, Gil L, Jupp S, et al. The polygenic score catalog as an open database for reproducibility and systematic evaluation. Nat Genet 2021;53:420–5. 10.1038/s41588-021-00783-5

21 The Polygenic Score (PGS) Catalog. Available: https://www.pgscatalog.org/

22 Wald NJ. Guidance on terminology. J Med Screen 2023;30:53–4. 10.1177/09691413231163128

23 Wald NJ, Morris JK. Assessing risk factors as potential screening tests: a simple assessment tool. Arch Intern Med 2011;171:286–91. 10.1001/archinternmed.2010.378

24 Wald NJ, Bestwick JP. Is the area under an ROC curve a valid measure of the performance of a screening or diagnostic test. J Med Screen 2014;21:51–6. 10.1177/0969141313517497

25 Wald N, Leck I. Antenatal and neonatal screening. 2000. 10.1093/acprof:oso/9780192628268.001.0001

26 Sun L, Pennells L, Kaptoge S, et al. Polygenic risk scores in cardiovascular risk prediction: a cohort study and modelling analyses. PLoS Med 2021;18:e1003498. 10.1371/journal.pmed.1003498

27 NICE. Overview | Cardiovascular disease: risk assessment and reduction, including lipid modification | Guidance, Available: https://www.nice.org.uk/guidance/cg181 [Accessed 5 Jan 2022].

28 Wald NJ, Simmonds M, Morris JK. Screening for future cardiovascular disease using age alone compared with multiple risk factors and age. PLoS One 2011;6:e18742. 10.1371/journal.pone.0018742

29 Wald NJ, Law MR. A strategy to reduce cardiovascular disease by more than 80. BMJ 2003;326:1419. 10.1136/bmj.326.7404.1419

30 Breast cancer risk in American women - NCI. Available: https://www.cancer.gov/types/breast/risk-fact-sheet [Accessed 25 Aug 2022].

31 Elmore JG, Armstrong K, Lehman CD. Screening for breast cancer. JAMA 2005;293:1245. 10.1001/jama.293.10.1245

32 Law MR, Wald NJ. Risk factor thresholds: their existence under scrutiny. BMJ 2002;324:1570–6. 10.1136/bmj.324.7353.1570

33 Rose G. Sick individuals and sick populations. Int J Epidemiol 1985;14:32–8. 10.1093/ije/14.1.32

34 Holmes MV, Harrison S, Talmud PJ, et al. Utility of genetic determinants of lipids and cardiovascular events in assessing risk. Nat Rev Cardiol 2011;8:207–21. 10.1038/nrcardio.2011.6

35 Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test BMJ 1999;319:1562–5. 10.1136/bmj.319.7224.1562

36 Hingorani AD, Hemingway H. How should we balance individual and population benefits of statins for preventing cardiovascular disease. BMJ 2010;342:c6244. 10.1136/bmj.c6244

37 Winyard G. SMAC statement on the use of statins. London: Department of Health, 1997.

38 Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. J Am Coll Cardiol 2019;74:1376–414. 10.1016/j.jacc.2019.03.009

39 Hingorani AD, Windt DA van der, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ 2013;346:e5793. 10.1136/bmj.e5793

40 Huntley C, Torr B, Sud A, et al. Utility of polygenic risk scores in UK cancer screening: a modelling analysis. Lancet Oncol 2023;24:658–68. 10.1016/S1470-2045(23)00156-0

41 Polygenic risk scores - Genomics Plc. Available: https://www.genomicsplc.com/resources/polygenic-risk-scores/ [Accessed 11 Nov 2021].

42 Direct-to-consumer genomic testing: government response to the committee's first report - science and technology committee - house of commons. Available: https://publications.parliament.uk/pa/cm5802/cmselect/cmsctech/805/80502.htm [Accessed 10 Jan 2022].

43 RSS - working group on diagnostic tests. Available: https://rss.org.uk/policy-campaigns/policy-groups/working-group-on-diagnostic-tests/ [Accessed 26 Jun 2023].

44 Schmidt AF, Hingorani AD, Finan C. Human genomics and drug development. Cold Spring Harb Perspect Med 2022;12:a039230. 10.1101/cshperspect.a039230

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjmed-2023-000554).

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJMED

Performance of polygenic risk scores in screening, prediction, and risk stratification – Extended Methods

**Hingorani AD., FMedSci[1,2,3,4], Gratton J., PhD[1, 2], Finan C., PhD[1,2,3,4], Schmidt AF.,
PhD[1,2,3,4,5], Patel R., PhD[1,2,3,4]., Sofat R., PhD[3,4,6], Kuan V., PhD[1,2,3], Langenberg C., PhD[7],
Hemingway H., FMedSci[2,3,4,8], Morris JK., PhD[9], Wald NJ., FRS[8,9]**

[1] Institute of Cardiovascular Science, University College London, UK.

[2] UCL British Heart Foundation Research Accelerator.

[3] UCL National Institute of Health Research Biomedical Research Centre.

[4] Health Data Research, UK.

[5] University Medical Centre, Utrecht, Netherlands.

[6] Department of Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool.

[7] Queen Mary University of London, UK.

[8] Institute of Health Informatics, University College London, UK.

[9] Population Health Research Institute, St. George's, University of London, UK.

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJMED

Observed characteristics of polygenic scores are that they exhibit a Gaussian distribution, with the distribution being higher among those affected than among those unaffected, with the same SD in both groups. Consequently, the proportional difference in disease risk is the same for any given absolute difference in polygenic score value, from any starting level (i.e., a log-linear relationship). These properties enable calculation of metrics useful in the evaluation of performance is screening, individual risk prediction and population risk stratification.

Deriving DR_5 from HR or OR per-SD

The detection rate (DR) for a given false positive rate can be derived from the OR or HR per-SD. The HR is the relative risk of the event occurring at any point in time compared to a control group of people and therefore all the assumptions for the OR are also applicable for the HR. Consider the Gaussian frequency distributions of polygenic risk scores among those unaffected (U) and those who become affected by disease (A). The mean polygenic score value among those unaffected is μ_U and among those affected is μ_A, and both distributions have the same SD (σ)
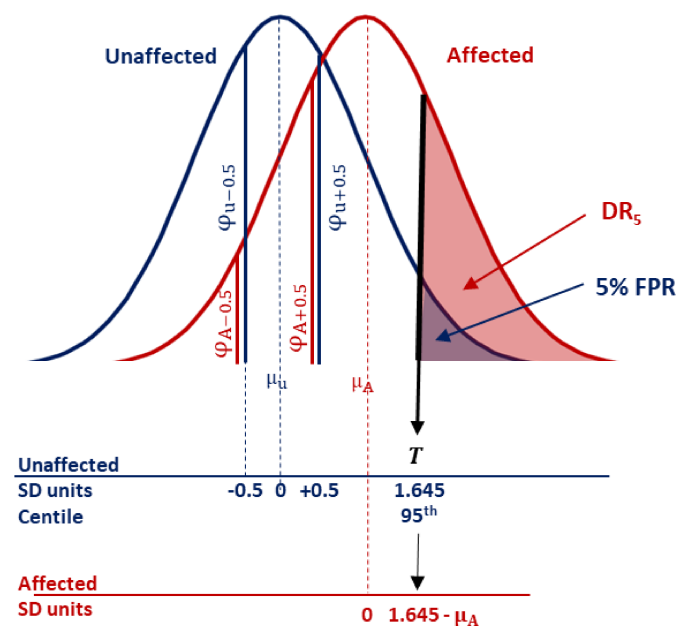
The OR per-SD can be considered to be equal to the ratios of the probability density function values φ, for affected (φ_A) vs. unaffected individuals (φ_U) at Z-score values for the distribution among the unaffected (i.e. Z_U values) that are 1-SD apart, for example

corresponding to values at $+0.5\ SD$ and $-0.5 SD$. In the figure, the values are shown by vertical red and blue, lines respectively. Thus,

$$OR_{SD} = \frac{\left(\varphi_A/\varphi_U\right)_{+0.5}}{\left(\varphi_A/\varphi_U\right)_{-0.5}}$$

Where x is the value of the variate, the probability density function φ is given by:

$$\varphi(\mu, \sigma, x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

For the distribution in the unaffected, $\mu = 0$, $\sigma = 1$

$$\varphi U(0,1,x) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{1}\right)^2}$$

$$\varphi U(0,1,-0.5) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{-0.5}{1}\right)^2}$$

$$\varphi U(0,1,+0.5) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{+0.5}{1}\right)^2}$$

Since $\varphi_{U+0.5} = \varphi_{U-0.5}$,

$$OR_{SD} = \frac{\left(\varphi_{A+0.5}\right)}{\left(\varphi_{A-0.5}\right)}$$

For the distribution in the affected, $\mu = \mu_A$, $\sigma = 1$

$$\varphi A(\mu_A, \sigma, -0.5) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{-\mu_A - 0.5}{1}\right)^2}$$

$$\varphi A(\mu_A, \sigma, +0.5) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{0.5 - \mu_A}{1}\right)^2}$$

Thus,

$$OR_{SD} = \frac{\frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{0.5-\mu_A}{1}\right)^2}}{\frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{-\mu_A-0.5}{1}\right)^2}}$$

Cancelling $\frac{1}{1\sqrt{2\pi}}$,

$$OR_{SD} = \frac{e^{-\frac{1}{2}(0.5-\mu_A)^2}}{e^{-\frac{1}{2}(-\mu_A-0.5)^2}}$$

$$OR_{SD} = e^{\left[-\frac{1}{2}(0.5-\mu_A)^2 + \frac{1}{2}(-\mu_A-0.5)^2\right]}$$

$$OR_{SD} = e^{-\frac{1}{2}\left[(0.5-\mu_A)^2 - (-\mu_A-0.5)^2\right]}$$

$$OR_{SD} = e^{-\frac{1}{2}[-2\mu_A]}$$

$$OR_{SD} = e^{\mu_A}$$

$$\mu_A = \ln OR_{SD}$$

Also from the figure,

$$DR_5 = 1 - \emptyset(1.645 - \mu_A)$$

where $\emptyset$ corresponds to the cumulative distribution function (CDF) corresponding to the area under the normal distribution among affecteds at a Z-value of $1.645 - \mu_A$. (The 1.645 is the 95th centile of the distribution among unaffecteds, which is the threshold for a 5% false positive rate). Because of the symmetry of the normal distribution,

$$DR_5 = \emptyset(\mu_A - 1.645)$$

Thus, from equation 2,

$$DR_5 = \emptyset(\ln OR_{SD} - 1.645)$$

Consider, as an example, an $OR_{SD} = 5$.

In this example, $\ln OR_{SD} = \mu_A = 1.609$.

Thus, from equation 3,
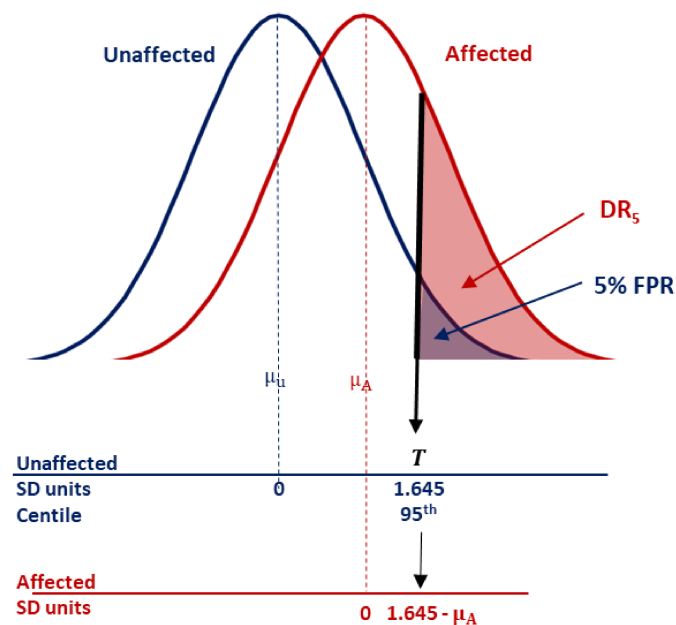
$$DR_5 = \emptyset(1.609 - 1.645)$$

$$DR_5 = \emptyset(-0.03556)$$

$$DR_5 = 0.485816$$

Thus an OR per-SD of 5 corresponds to a 49% detection rate for a 5% false positive rate.

Estimating DR_5 from AUC and C-index

The following derivation is adopted from Wald and Bestwick (Ref 23).

The DR at a cut-point value T, is given by,

$$DR = 1 - \emptyset\left(\frac{T - \mu_A}{\sigma_A}\right)$$

Because of the symmetry of the normal distribution,

$$DR = \emptyset\left(\frac{\mu_A - T}{\sigma_A}\right)$$

By the same argument,

$$FPR = \emptyset\left(\frac{\mu_U - T}{\sigma_U}\right)$$

The Z-score corresponding to the DR expressed in terms of the affected distribution is:

$$\emptyset^{-1}(DR) = \frac{\mu_A - T}{\sigma_A}$$

The Z-score corresponding to the FPR expressed in terms of the unaffected distribution is:

$$\emptyset^{-1}(FPR) = \frac{\mu_U - T}{\sigma_U}$$

Writing both equations in terms of T

$$T = \mu_A - \sigma_A\emptyset^{-1}(DR)$$

$$T = \mu_U - \sigma_U\emptyset^{-1}(FPR)$$

Thus,

$$\mu_A - \sigma_A\emptyset^{-1}(DR) = \mu_U - \sigma_U\emptyset^{-1}(FPR)$$

Making DR the subject,

$$\sigma_A\emptyset^{-1}(DR) = \mu_A - \mu_U + \sigma_U\emptyset^{-1}(FPR)$$

$$\emptyset^{-1}(DR) = \frac{\mu_A - \mu_U + \sigma_U\emptyset^{-1}(FPR)}{\sigma_A}$$

$$DR = \emptyset\left(\frac{\mu_A - \mu_U}{\sigma_A} + \frac{\sigma_U\emptyset^{-1}(FPR)}{\sigma_A}\right)$$

Considering the standard normal distribution and assuming equal standard deviation in unaffected and affected groups, and $\mu_U = 0$.

$$DR = \emptyset \left(\frac{\mu_A}{1} + \frac{1 \cdot \emptyset^{-1}(FPR)}{1} \right)$$

$$DR = \emptyset \left(\mu_A + \emptyset^{-1}(FPR) \right)$$

The area under the ROC curve (AUC), or equivalently the C-index, is the probability that an affected individual drawn at random (A) has a higher polygenic risk score than an unaffected individual drawn at random (U) i.e.

$$P(A > U) = P(A - U > 0)$$

The AUC is the therefore the CDF for the distribution of differences (the variances sum). Thus,

$$AUC = \emptyset \left(\frac{\mu_A - \mu_U}{\sqrt{\sigma_A^2 + \sigma_U^2}} \right)$$

$$\emptyset^{-1}(AUC) = \frac{\mu_A - \mu_U}{\sqrt{\sigma_A^2 + \sigma_U^2}}$$

Given that for the standard normal distribution $\mu_U = 0$ and if we assume equal standard deviation for the distributions for affected and unaffected individuals, $\sigma_U = \sigma_A = 1$,

$$\emptyset^{-1}(AUC) = \frac{\mu_A}{\sqrt{2}}$$

For example, if the AUC is 0.7,

$$\emptyset^{-1}(0.7) = 0.524$$

Thus,

$$\frac{\mu_A}{\sqrt{2}} = 0.524$$

$$\mu_A = 0.524 \cdot \sqrt{2}$$

$$\mu_A = 0.742$$

Since,

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJMED

$$DR = \emptyset\big(\mu_A + \emptyset^{-1}(FPR)\big)$$

the DR_5 corresponding to an AUC of 0.7 is,

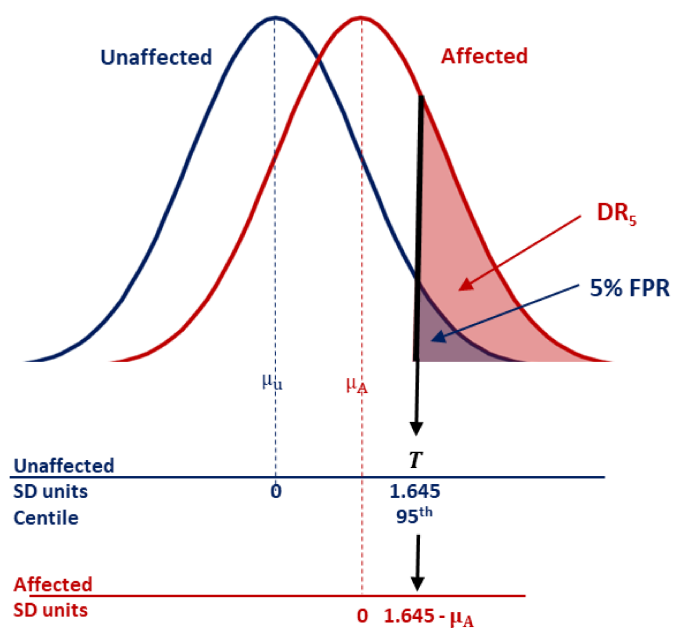$$DR_5 = \emptyset\big(0.742 + \emptyset^{-1}(0.05)\big)$$

$$DR_5 = \emptyset(0.742 + (-1.645)$$

$$DR_5 = \emptyset(-0.903)$$

$$DR_5 = 0.183$$

Thus an AUC of 0.7 corresponds to a detection rate of 18% for a 5% false positive rate.

Calculating the likelihood ratio and odds of becoming affected given a positive test result (OAPR)

In evaluating the performance of a polygenic risk score as a screening test, we calculated the likelihood ratio for a positive result (i.e., a polygenic risk score at or above a pre-specified

cut-off) as the ratio DR/FPR. The likelihood ratio for a test cut-off with a 5% FPR is given by $DR_5/_{5\%}$, where DR_5 is expressed as a percentage. The $OAPR$ is calculated by multiplying the background odds of disease by the likelihood ratio for a positive test result. For example, if the background odds of disease in the population is $1:9$ and the DR_5 is 15%, the likelihood ratio is $15/5 = 3$ and the $OAPR = 3:9$ or $1:3$.

Calculating the likelihood ratio and odds of becoming affected for an individual with a given polygenic score result
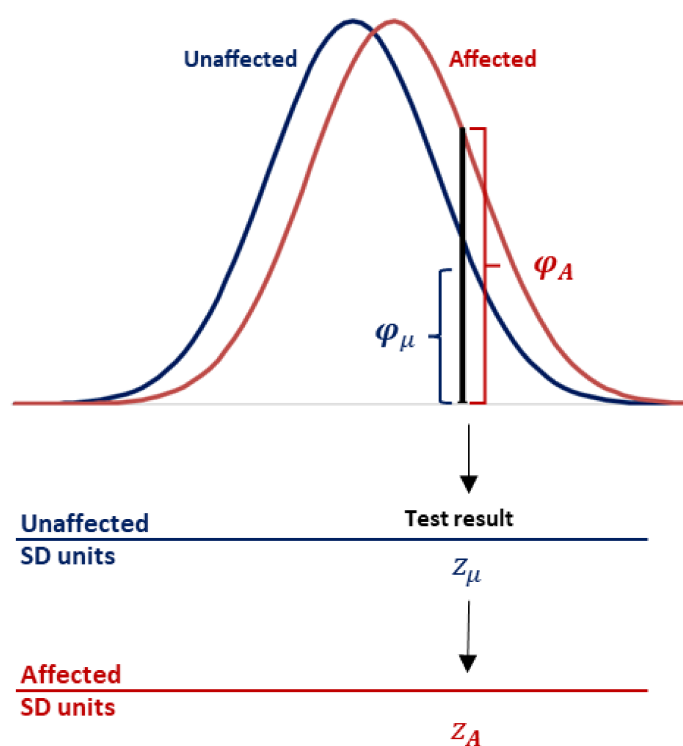
To compute how many times more likely a given polygenic score result is to arise from an affected than unaffected individual, we compared the heights of the standard distribution curves for affected and unaffected individuals at that value (the likelihood ratio; LR). In the illustration shown, this is given by the ratio $\varphi_A/_{\varphi_U}$.

$$LR = \varphi_A/_{\varphi_U}$$

These heights can be calculated using the equation for the Gaussian distribution:

$$LR = \frac{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(Z_A)^2}}{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(Z_U)^2}}$$

For example, for a polygenic risk score with a performance metric expressed as an $OR_{SD} = 1.61$:

$$\mu_A = \ln OR_{SD} = 0.48$$

A polygenic score at the 75th centile of the distribution for unaffected individuals yields,

$Z_U = 0.67; \varphi_U = 0.32$

$Z_A = 0.2, \varphi_A = 0.39$

$LR = \varphi_A/\varphi_U = 0.39/0.32 = 1.23$

If the background odds of disease in the population is $1:9$, an individual whose polygenic score is at the 75th centile of the distribution among unaffecteds has an odds of being affected of $(1.23 \times 1):9 \approx 1:7$.

Calculating the likelihood ratio and odds of becoming affected for a particular polygenic risk score group
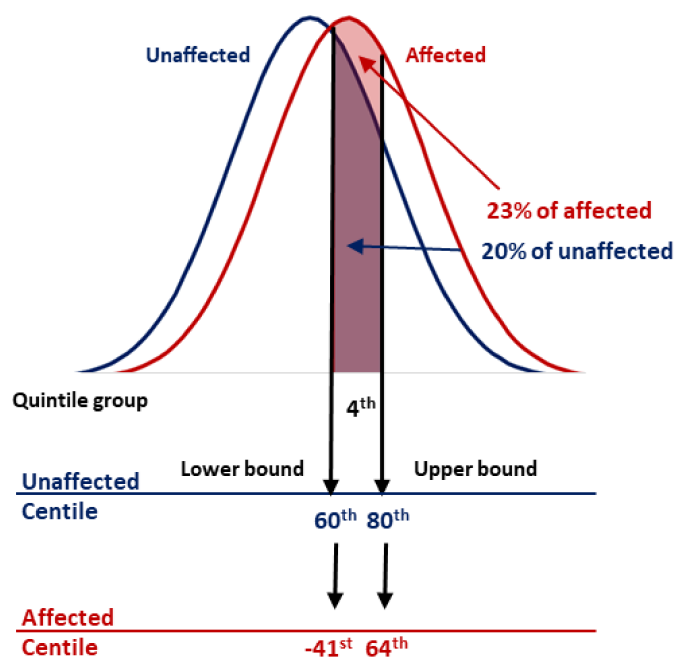
In evaluating polygenic risk scores in risk stratification, we calculated the LR as the ratio of areas under the distributions for affected and unaffected individuals in each polygenic score quantile (e.g., each fifth of the polygenic score distribution with respect to the unaffected, as shown in the figure). We then multiplied the background odds of disease by the corresponding likelihood ratio to determine the odds of becoming affected for each quantile of the distribution. For example, for individuals in the fourth quintile, if the LR is 1.2 (figure) and the background odds of disease in the population is $1:9$, the odds of becoming affected for this group is $1.2:9 \approx 1:8$.