medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Performance of polygenic risk scores in screening, prediction, and risk stratification

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

Corresponding author: Aroon Hingorani (a.hingorani@ucl.ac.uk)

Keywords

Polygenic risk score; Screening; Prediction; Risk Stratification

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background: The clinical value of polygenic risk scores has been questioned. We sought to clarify performance in population screening, individual risk prediction and population risk stratification by analysing 926 polygenic risk scores for 310 diseases from the Polygenic Score (PGS) Catalog.

Methods: Polygenic risk scores in the PGS Catalog are reported using hazard ratios or odds ratios per standard deviation, or the area under the receiver operating characteristic curve sometimes expressed as the C-index. We used this information to produce estimates of performance in: (a) *population screening* — by calculating the detection rate (DR_5) for a 5% false positive rate (FPR) and the population odds of becoming affected given a positive result $(OAPR)$; (b) *individual risk prediction* — by calculating the individual odds of becoming affected for a person with a particular polygenic score; and (c) *population risk stratification* — by calculating the odds of becoming affected for groups of individuals in different portions of a polygenic risk score distribution. We use coronary artery disease and breast cancer as illustrative examples.

Findings: *Population screening performance*: The median DR_5 for all polygenic risk scores and all diseases studied was 11% [interquartile range $8-18\%$]. The median DR_5 was 12% $[9-19]$ for polygenic risk scores for CAD and 10% $[9-12]$ for breast cancer, with population $OAPRs$ of $1:8$ and $1:21$ respectively, with background 10-year odds of $1:19$ and $1:41$ respectively, which are typical for these diseases at age 50. *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 $97.5th$ centile were $1:54$, $1:29$, $1:15$, and $1:8$ for CAD and $1:91$, $1:56$, $1:34$, and $1:21$ for breast cancer. *Population risk stratification*: At age 50, stratifying into quintile groups of CAD risk yielded 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 CAD, a group that contributed 7% of cases. The corresponding estimates for breast cancer were $1:72$ and $1:26$ for lowest and highest quintiles; and $1:19$ for the upper 2.5% of the distribution, which contributed 6% of cases.

Interpretation: Polygenic risk scores perform poorly in population screening, individual risk prediction, and population risk stratification.

Funding: British Heart Foundation; UK Research and Innovation; National Institute of Health and Care Research.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

A polygenic risk score represents the weighted sum of independent DNA sequence variants that are present in an individual's genome which increase 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 (GWAS). The increasing range and scale of GWAS over the last decade, now spanning over 2500 diseases or traits[2], has led to a proliferation in polygenic risk scores, igniting widespread interest in potential healthcare applications[3], and capturing the attention of 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 are ready to implement in practice[7]. In a progression toward clinical implementation, position papers have appeared on reporting standards and responsible clinical use from the Clinical Genome Resource (ClinGen) Complex Disease Working Group[8], and the Polygenic Risk Score Task Force of the International Common Disease Alliance[9]. Individual consumers and healthcare providers can already access commercial genetic testing and software services based on polygenic scores[10–12]. A 'world-first pilot' trial of predictive genetic testing for cardiovascular disease is also underway in participants attending Vascular Health Checks in the NHS[13], and 'genetic risk scores' for disease prediction are central to aims of the Our Future Health programme aiming to recruit 5 million UK adults[14].

These claims, however, are disputed. There is disagreement on the performance of polygenic risk scores in population screening, individual risk prediction and population risk stratification which makes their eventual role in medicine and public health uncertain[15–18]. Recently, Lambert and colleagues produced the PGS Catalog, a comprehensive, regularly updated, open-access directory of studies on *polygenic scores* for quantitative traits (e.g. blood pressure) and *polygenic risk scores* for diseases (e.g. breast cancer)[19]. The catalogue lists the following 'performance metrics' for polygenic risk scores: the hazard (HR) or odds ratio (OR), both per 1-standard deviation (SD) increment in the score, or the area under the receiver operating characteristic curve (AUC), sometimes expressed as the C-index. However, these metrics are not directly informative of performance in population screening, individual risk prediction, or population risk stratification. An appropriate metric 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 1 in 10 or 10%. The odds of becoming affected is obtained by multiplying the background odds of developing disease in a specified time frame by the likelihood ratio (LR) associated with a 'positive test' (in the case of population screening), with a particular polygenic score value (in the case of individual risk prediction),

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

or occupancy of a particular polygenic risk score quantile group (in the case of population risk stratification).

We mathematically derive the odds of becoming affected in each of these scenarios using the metrics reported in the PGS Catalog. We use breast cancer and coronary artery disease as illustrative examples and scrutinise two proposed early clinical uses of polygenic risk scores: to to improve on the performance of the established risk factor models in the prediction of CAD and stroke, and to prioritise mammographic screening at a younger age for the detection of breast cancer.

Methods

By April 2022, the PGS Catalog had curated $13\,828$ performance metric estimates for 2194 polygenic scores ('unique polygenic score codes'), involving 544 diseases or traits ('unique experimental factor ontology identifiers'), reported in 303 unique publications. We removed polygenic scores for continuous traits, and polygenic risk scores with implausible values (167 instances where the HR or OR per-SD was recorded as < 1, two instances where the AUC was < 0.5, and one instance where the C-index was recorded as 632), leaving 3915 performance metric estimates for 926 polygenic risk scores involving 310 unique binary outcomes (mainly diseases). The reported performance metrics were OR per-SD in 1216 instances, HR per-SD in 378 instances, AUC in 2077 and C-index in 244 instances (**Supplementary Tables)**.

Polygenic risk scores display a Gaussian distribution, so it is possible to use the originally reported metrics in the PGS Catalog to calculate the difference in mean values between individuals who are later affected or remain unaffected by disease (**Supplementary Tables**).[20,21] We used the difference in mean values to estimate the overlap in Gaussian distributions and calculate the DR and FPR which are the percentage of people with a polygenic score above a particular cut-off ('a positive test') among those who are later affected or remain unaffected by disease, respectively. For simplicity and consistency, we set polygenic score cut-offs that define a 5% FPR and calculate the corresponding detection rate (DR_5) [20,22]. However, we also provide estimates for DR_{10} and DR_1, the detection rates at 10% and 1% false positive rates (**Supplementary Tables**). Detection rates at any other false positive rate can be derived by adapting the equations in the **Extended Methods**. We defined the LR in screening as the ratio DR/FPR. In individual risk prediction, we defined LR as the ratio of the heights of the Gaussian distribution curves for affected and unaffected individuals at a particular polygenic risk score centile; and in risk stratification as the ratio of

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

areas under the relative frequency distributions for affected and unaffected individuals in each polygenic score quantile (e.g., each fifth, or the upper 2.5% of the polygenic score distribution, see **Extended Methods**). In each case, multiplying the LR by the background odds of disease for the whole population gives the corresponding odds of becoming affected for the individual or group of interest. Where we refer to a polygenic score centile or quantile, this is in reference to the distribution in the unaffected group. When referring to a particular polygenic risk score, we used the PGS Catalog identifier number.

We re-analysed data from two original sources[5,23] to quantify the extent to which the addition of polygenic risk score information to conventional risk factors improves the prediction of CAD and stroke. To do this, we calculated the DR and FPR with and without polygenic risk score information, using 10-year risk cut-offs recommended in guidelines for the initiation of statin treatment. We then calculated the number of individuals who need to be genotyped (and a polygenic score calculated) to detect or prevent one additional CAD event or stroke. We refer to this value as the 'number-needed-to-genotype'. We modelled the use of a breast cancer polygenic risk score to prioritise mammographic screening at age 40 rather than from the currently recommended 50 years of age.

Role of the funding source

None of the study sponsors played a role in study design; in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication

Results

Performance of polygenic risk scores in screening

For all diseases studied, the median DR_5 based on all polygenic risk scores was 11%, interquartile range $[8-18\%]$, i.e., 89% $[82-92]$ of cases missed. The median DR_5 values for polygenic risk scores whose performance was reported using OR or HR per-SD were 9% $[6-12]$ and 8% $[7-10]$ respectively. For polygenic risk score performance reported using AUC or C-index, the median DR_5 values were 14% $[10-22]$ and 19% $[13-25]$ respectively. The median DR_5 values for polygenic risk scores for 28 common diseases, including CAD and breast cancer are shown in **Figure 1**.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Coronary artery disease: The median DR_5 from performance metrics for 27 polygenic risk scores for CAD was 12% $[9-20]$ (88% of cases missed, **Figure 2**), which corresponds to an LR of 2.4. Applied at age 50, with a background 10-year CAD risk of 5% (odds $1:19$), the $OAPR = (2.4:19) \approx 1:8$, i.e., false positives outnumber true positives by around eight to one. Reducing the cut-off to reduce the FPR to 1% reduces the DR to 3%, with 97% of cases missed. Retaining a 5% FPR but applying the test in a population with 2% CAD incidence over the same period (background odds $1:55$), e.g., at around 40 years of age, yields an $OAPR$ of $1:23$, with false positives outnumbering true positives by just over twenty to one.

Breast cancer: The median DR_5 from performance metrics reported for 108 polygenic risk scores for breast cancer was 10% $[9-12]$ (90% of cases missed **Figure 2**). This corresponds to an LR of 2. Applied at age 50, with a background 10-year breast cancer risk of about 2.5% (odds $1:41$), the $OAPR = 1:21$. Applying the polygenic risk score as a test at age 40, when the background 10-year odds is $1:64$ yields an $OAPR$ of $1:32$, false positives outnumbering true positives by just over thirty to one.

Performance of polygenic risk scores in individual risk prediction

The overlap in polygenic score distributions derived from the metrics in the PGS Catalog enables calculation of the LR 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* (see **Extended Methods**)[22].

Coronary artery disease: The odds of developing CAD 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 $97.5th$ centile respectively at age 50 (when the background odds is $1:19$) (**Figure 3**); and $1:157$, $1:85$, $1:45$, and $1:24$ respectively at age 40 (when the background odds is $1:55$).

Breast cancer: The average 10-year odds of breast cancer is $1:41$ 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 3)**, 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 $97.5th$ centile respectively.

Performance of polygenic risk scores in risk stratification

Coronary artery disease: The left panel in **Figure 4** shows the overlapping distributions for affected and unaffected individuals applied to a hypothetical cohort of $100{,}000$ 50-year-old men with stratification into polygenic risk score quintile groups. The 10-year odds of CAD are

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

reduced from the average of $1:19$ for all men to $1:41$ for those in the lowest quintile group and increased to $1:11$ in the highest quintile group. In **Figure 5**, focusing on those at the highest risk, the 10-year odds was $1:7$ for the upper 2.5% of the polygenic risk score distribution for CAD, but this group contributed only 7% of cases (**Figure 5**).

Breast cancer: In the right-hand panel of **Figure 4**, we applied the same approach to a hypothetical cohort of 50-year-old women with a background 10-year odds of breast cancer of $1:41$. Odds reduce by about half to $1:72$ for those in the lowest quintile and almost double to about $1:26$ for those in the highest quintile of the polygenic risk score distribution. For the highest 2.5% of the polygenic risk score distribution in the right-hand panel of **Figure 5**, odds are increased to $1:19$, but the latter group only accounts for 6% of breast cancer cases.

Screening using polygenic risk scores together with conventional risk factors or tests

i) CAD

It has been proposed that adding polygenic risk scores to conventional risk factors (e.g., blood pressure and LDL-cholesterol), would usefully improve CAD and stroke screening to indicate who should be offered a statin prescription for primary prevention. The **Table** shows the results from Sun *et al*.[23], applied in a hypothetical cohort of 100,000 40-year old individuals with a risk factor profile representative of the English population and a background 10-year CAD and stroke risk of 8%. A conventional, multi-risk factor model incorporating age and using a 10% 10-year risk cut off, detected 60% of those later affected by CAD or stroke at a 24% FPR ($DR_{24} = 60\%$). The addition of polygenic risk scores for CAD and stroke to the model (PGS Catalog identifiers PGS000018 and PGS000039 respectively) led to the detection of 61% of those affected for a 23% FPR ($DR_{23} = 61\%$). Assuming a 10-year risk cut-off of 10% for prescribing statins[24], 100% adherence, and adopting the assumption of Sun *et al*. that statins reduce the risk of CAD and stroke by 20%[23] , 974 events would be prevented using a model based on conventional risk factors *and* polygenic risk scores instead of 957 using a conventional risk factor model with no genetic information, a gain of 17 cases. This gives a number needed-to-genotype to prevent one additional event of 5882 (**Table**). Sun *et al*. also estimated that 1029 CAD and stroke events would be prevented using a hybrid model, where conventional risk factor assessment is followed by polygenic risk scores only for those at intermediate ($5 - 10\%$) 10-year risk. However, replacing this more complicated model with one where the whole cohort receives statins would prevent 1600 cardiovascular events using the same assumptions (**Supplementary Tables**). Since age is the major determinant of CAD and stroke risk, age alone performs about as well as muti factor risk models that include age[25]. Given the rarity

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

of CAD and stroke events below 50, using an age cut-off of 50 instead of 40 would prevent almost as many events but with many fewer false positives[26].

Similar results were obtained using data reported by Riveros-McKay et al.[5]. These authors also investigated the extent to which the addition of a polygenic risk score to conventional risk factors improves the identification of UK Biobank participants eligible to receive statins because their 10-year risk of CAD and stroke exceeds the cut-offs used in UK or US primary prevention guidelines. Deriving the appropriate metrics from their data (**Table and Supplementary Tables**) clarifies the effect of adding information from a polygenic risk score for CAD. Using a 10-year risk cut-off of 10% for initiating statins, the QRISK3 model based on conventional risk factors including age detected 81% of cases at a 42% FPR ($DR_{42} = 81\%$). The addition of a polygenic risk score to the model detected 84% of cases for a 41% FPR ($DR_{41} = 84\%$). Using the authors' assumption that statins reduce CAD and stroke events by 20%, the number needed-to-genotype to prevent one additional event based on this study is 8879 (**Table**).

ii) Breast cancer

It has also been proposed that polygenic risk scores are used to prioritise the use of established screening tests for cancer[3]. One suggestion is that younger women should undergo mammographic screening if their breast cancer risk determined using a polygenic risk score exceeds that of an average 50-year-old, the age beyond which mammography is offered to all women. **Figure 5** shows that 40-year-old women 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 $1:41$ [27]. **Figure 6** shows that using the breast cancer polygenic risk score (PGS Catalog identifier PGS000004) as a stage 1 screen in 100,000 40-year-old women, applying the unaffected $97.5th$ centile as a cut-off, would result in 2570 women with a 'high-risk' polygenic score being offered mammography of whom 108 would be affected and 2462 unaffected ($OAPR$ $1:23$). Assuming 100% uptake and a DR_8 of 75%[28], mammography would then correctly identify 81 of the 108 affected individuals but miss 27 breast cancer cases. However, 1430 breast cancer cases (over ten times as many) are estimated to occur among the $97\ 430$ 40-year-old women with a polygenic risk score below the unaffected $97.5th$ centile who would not be offered mammography.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Discussion

Our results show the poor performance of polygenic risk scores in population screening, individual disease prediction, and population risk stratification. This is not obvious from the metrics reported in the PGS Catalog but is clear using the appropriate metrics employed in this study. Our conclusion is consistent with that of other authors,[17,18,29] but is insufficiently recognised. The findings are relevant to consumers, patients, doctors, those involved in preventative medicine and public health, as well as funders and policy makers.

Polygenic risk score distributions overlap substantially for all conditions studied and this extensive overlap constrains their performance in each of their intended applications, whether used alone or in conjunction with conventional risk factors or screening tests. For instance, 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 OR per-SD of 12 or higher (compared to the median observed value of 1.31) or an AUC of 0.96 (compared to the median observed value of 0.65). Only 11.4% of AUC values in the PGS 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 **Supplementary Tables**).

Where a risk factor displays a monotonic relationship with disease risk[30], more cases arise among the majority with near average risk factor values than among the few with more extreme values – the 'prevention paradox' [31 32]. In this respect, polygenic risk scores are similar to certain non-genetic risk factors such as blood pressure and LDL-cholesterol which, though causal, are poor predictors of CAD[17,33]. That the performance of polygenic risk scores in the prediction of CAD is sometimes compared favourably to that of blood pressure and cholesterol[23] is simply to benchmark one poor predictor against another.

Where there are safe and inexpensive preventative interventions (e.g., statins and blood pressure lowering drugs for prevention of CAD and stroke) there is greater public health benefit in broadening rather than limiting eligibility for such interventions[34]. In the prevention of CAD and stroke, this has been achieved *de facto* by the progressive lowering of the 10-year risk cut-off for statin prescription in primary prevention. The cut-off was reduced from a 30% 10-year risk of CAD in the UK in 1997[35], down to 10% 10-year risk of CAD or stroke in the UK from 2016[24] and 7.5% in the US from 2019[36]. The reduction in the risk cut-off was enabled by falling drug acquisition cost through patent expiry, and by accumulating evidence on long-term safety. Eligibility could be extended yet further and also simplified by using age alone to guide statin prescription for primary prevention with the prevention of many more CAD and stroke cases[25]. By contrast, retaining the same 10-year risk cut-off and adding polygenic risk score information to conventional risk factor models has a much weaker

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

impact. Using recently reported data[5,23], we show 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 (using genetics or other means) may be justified if a preventative intervention is costly, resource limited, or has significant harms[37]. However, as we demonstrated using breast cancer as an example, identifying those at high risk requires testing in all and, aside from missing the many more cases among those at average risk, generates many false positives. This could have substantial downstream resource implications for healthcare systems if, for example, genetic risk stratification were to be followed by a confirmatory screening test, such as mammography for breast cancer. In this case, it would be more sensible to simply reduce the age cut-off for mammography for all women without determining their polygenic risk score.

The enthusiasm surrounding polygenic risk scores may have been encouraged by pressure on academia to demonstrate a tangible health impact after decades of research investment in human genomics and by commercial opportunity. Unrealistic expectations have probably been raised by use of the wrong metrics. Publications on polygenic risk scores often illustrate comparisons between mutually exclusive groups, e.g., those in opposite tails of a polygenic score distribution[38]. This is relevant in aetiological studies but is not relevant in screening. As shown in **Figure 1**, seemingly impressive odds ratios of 13, 7, 5, 4 or 3, for comparisons of the top vs. the bottom 1%, 5%, 10%, 20% and 25% of the polygenic risk score distribution respectively, all equate to a DR_5 of only 12%. What is relevant in screening is the risk of an event in a group compared to that of the whole population, which is what calculation of the DR for a specified FPR achieves.

Our findings are relevant both 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, e.g., 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. In tandem, as already suggested[39], policy makers may wish to consider tighter regulation of commercial genetic tests based on polygenic risk scores, with a focus on clinical not just assay performance, 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: 1) the mean and SD of polygenic risk score values among later affected and unaffected individuals; 2) the overlap in their distributions; 3) the

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

relevant performance metrics such as the DR for a specified FPR, such as the DR_5, avoiding the need to calculate this indirectly[20]; and 4) the 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 the current analysis shows the poor performance of polygenic risk scores in screening, prediction, and risk stratification, they may find use in other situations. For example, polygenic scores may explain the variable penetrance of rare mutations in monogenic diseases, e.g., hypertrophic cardiomyopathy or familial hypercholesterolaemia, and be employed to aid case detection. There are also other predictive applications of genotyping, e.g., in pharmacogenetic testing, to optimise efficacy and safety of medicines. Genotyping may 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 is likely to be in providing evidence on disease causation and targets for pharmaceutical intervention[40].

In conclusion, use of the appropriate metrics has demonstrated the poor performance of polygenic risk scores in population screening, individual risk prediction, and population risk stratification. By virtue of its wide scope, the current study may help to resolve the debate on the value of polygenic risk scores and avoid unjustified expectations about their role in preventing disease.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Acknowledgement of funding

Supported by the UCL British Heart Foundation Accelerator (AA/18/6/34223), the UCL NIHR Biomedical Research Centre, and the UKRI/NIHR funded Multimorbidity Mechanism and Therapeutics Research Collaborative (MR/V033867/1). JG is a UCL British Heart Foundation PhD candidate (FS/17/70/33482). ADH and HH are NIHR Senior Investigators.

Declaration of interests

ADH is a member of the Advisory Group for the Industrial Strategy Challenge Fund Accelerating Detection of Disease Challenge, and a co-opted member of the National Institute for Health and Care Excellence Guideline update group for 'Cardiovascular disease: risk assessment and reduction, including lipid modification, CG181. ADH is a co-Investigator on a grant from Pfizer to identify potential therapeutic targets for heart failure using human genomics. NJW is a director of Polypill Limited, a company that provides an online cardiovascular disease prevention service accessed on Polypill.com

Data sharing

The data used in this manuscript are in the Supplementary Tables.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

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. DOI:10.1001/jamapsychiatry.2020.3049.

2 Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun. Biol. 2019. DOI: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. DOI:10.1038/s41576-018-0018-x.

4 Genome UK: the future of healthcare - GOV.UK. https://www.gov.uk/government/publications/genome-uk-the-future-of-healthcare (accessed Jan 24, 2022).

5 Riveros-Mckay F, Weale ME, Moore R, *et al.* Integrated Polygenic Tool Substantially Enhances Coronary Artery Disease Prediction. *Circ Genomic Precis Med* 2021. DOI:10.1161/CIRCGEN.120.003304.

6 For health systems - Genomics plc. https://www.genomicsplc.com/precision-health/health-systems/ (accessed Aug 25, 2022).

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

8 Wand H, Lambert SA, Tamburro C, *et al.* Improving reporting standards for polygenic scores in risk prediction studies. *Nature* 2021; **591**: 211–9.

9 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 Horton R, Crawford G, Freeman L, Fenwick A, Wright CF, Lucassen A. Direct-to-consumer genetic testing. *BMJ* 2019. DOI:10.1136/bmj.l5688.

11 Genomics PLC - Genomics plc. https://www.genomicsplc.com/ (accessed June 5, 2022).

12 Allelica | Polygenic Risk Score. https://www.allelica.com/ (accessed June 5, 2022).

13 NHS pilots genetic testing to predict risk of heart disease | Heart disease | The Guardian. https://www.theguardian.com/society/2022/jan/17/nhs-pilots-genetic-testing-to-predict-risk-of-heart-disease (accessed Sept 11, 2022).

14 Health of nation study calls on millions to sign up - BBC News. https://www.bbc.co.uk/news/health-63331805 (accessed Oct 25, 2022).

15 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 2018. DOI:10.1038/s41576-018-0018-x.

16 Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Hum. Mol. Genet. 2019. DOI:10.1093/hmg/ddz187.

17 Wald NJ, Old R. The illusion of polygenic disease risk prediction. Genet. Med. 2019. DOI:10.1038/s41436-018-0418-5.

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

19 Lambert SA, Gil L, Jupp S, *et al.* The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat. Genet. 2021. DOI:10.1038/s41588-021-00783-5.

20 Wald NJ, Morris JK. Assessing risk factors as potential screening tests: A simple assessment tool. *Arch Intern Med* 2011. DOI:10.1001/archinternmed.2010.378.

21 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. DOI:10.1177/0969141313517497.

22 Wald N, Leck I. Antenatal and neonatal screening. 2009 DOI:10.1093/acprof:oso/9780192628268.001.0001.

23 Sun L, Pennells L, Kaptoge S, *et al.* Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. *PLoS Med* 2021. DOI:10.1371/JOURNAL.PMED.1003498.

24 Overview | Cardiovascular disease: risk assessment and reduction, including lipid modification | Guidance | NICE. https://www.nice.org.uk/guidance/cg181 (accessed Jan 5, 2022).

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

26 Wald NJ, Law MR. A strategy to reduce cardiovascular disease by more than 80%. *Br Med J* 2003. DOI:10.1136/bmj.326.7404.1419.

27 Breast Cancer Risk in American Women - NCI. https://www.cancer.gov/types/breast/risk-fact-sheet (accessed Aug 25, 2022).

28 Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. J. Am. Med. Assoc. 2005. DOI:10.1001/jama.293.10.1245.

29 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; published online Aug 22. DOI:10.1001/JAMAINTERNMED.2022.3171.

30 Law MR, Wald NJ. Risk factor thresholds: Their existence under scrutiny. Br. Med. J. 2002. DOI:10.1136/bmj.324.7353.1570.

31 Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985. DOI:10.1093/ije/14.1.32.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

32 Holmes M V., Harrison S, Talmud PJ, Hingorani AD, Humphries SE. Utility of genetic determinants of lipids and cardiovascular events in assessing risk. *Nat Rev Cardiol* 2011; **8**: 207–21.

33 Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? *BMJ* 1999. DOI:10.1136/bmj.319.7224.1562.

34 Hingorani AD, Hemingway H. How should we balance individual and population benefits of statins for preventing cardiovascular disease? *BMJ* 2011. DOI:10.1136/bmj.c6244.

35 Winyard G. SMAC statement on the use of statins. 1997.

36 Arnett DK, Blumenthal RS, Albert MA, *et al.* 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary. *J Am Coll Cardiol* 2019; **74**: 1376–414.

37 Hingorani AD, Van Der Windt DA, Riley RD, *et al.* Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ* 2013; **346**. DOI:10.1136/bmj.e5793.

38 Polygenic Risk Scores - Genomics plc. https://www.genomicsplc.com/resources/polygenic-risk-scores/ (accessed Nov 11, 2021).

39 Direct-to-consumer genomic testing: Government Response to the Committee's First Report - Science and Technology Committee - House of Commons. https://publications.parliament.uk/pa/cm5802/cmselect/cmsctech/805/80502.htm (accessed Jan 10, 2022).

40 Schmidt AF, Hingorani AD, Finan C. Human Genomics and Drug Development. *Cold Spring Harb Perspect Med* 2021. DOI:10.1101/cshperspect.a039230.

Original study	Risk tool	Screened	Genotyped	Risk cut-off	Detection rate	False positive rate	Number below cut-off	Events Below cut-off	Number above cut-off	Events above cut-off	Additional events detected	Events prevented	Additional events prevented	Number needed to genotype (detection)	Number needed to genotype (prevention)
Sun et al.	CRF	100000	0	≥10%	60%	24%	73277	3214	26722	4783	-	957	-	-	-
	CRF + PGS	100000	100000	≥10%	61%	23%	73554	3127	26445	4870	87	974	17	1149	5882
Riveros-Mckay et al	PCE	186451	0	>7.5%	74%	36%	118082	1112	68369	3135	-	627	-		-
	PCE+PGS	186451	186451	>7.5%	80%	36%	117516	855	68935	3392	257	678	51	725	3656
	QRISK3	186451	0	>10%	81%	42%	106697	797	79754	3450	-	690	-	-	-
	QRISK3+PGS	186451	186451	>10%	84%	41%	108359	690	78092	3557	107	711	21	1743	8879

Table. Effect of adding a polygenic risk score to non-genetic risk factors in prediction of coronary artery disease (CAD) and stroke. The values are based on data reported by Sun *et al*. (Ref 23) and Riveros-Mckay *et al*. (Ref 5). Both studies utilised data from UK Biobank. Sun *et al*. developed a conventional risk factor model (CRF) and examined the effect of adding polygenic risk scores for CAD (PGS000018) and stroke (PGS000039) on the prediction of subsequent CAD and stroke events. They used a 10-year risk cut-off of 10% for offering statin treatment. Riveros-McKay *et al*. modelled screening performance in 186,451 participants based on either the Pooled Cohort Equation (PCE) using a 7.5% 10-year risk cut-off, or QRISK3 using a 10% risk cut-off for statin prescription. The data on events reported by Riveros-Mackay *et al*. were for CAD alone rather than CAD and stroke. Calculations assume that all those exceeding the specified risk cut-off receive a statin, 100% adherence and that statin treatment produces a 20% relative risk reduction. Number needed to genotype refers to the number of individuals that need to be genotyped (and have a polygenic risk score calculated) to detect or prevent one additional cardiovascular event.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure legends

Figure 1. Performance in screening estimated for polygenic risk scores included in the PGS Catalog as of April 2022. The limits of each box represent the interquartile range and the horizontal line within each box is the estimated DR_5 value based on performance metrics reported for the corresponding polygenic risk scores. The selected diseases are colour coded into the following categories: cancers, cardiometabolic conditions, ocular diseases, allergic or autoimmune diseases, bone disease and neuropsychiatric diseases. The horizontal grey line is the estimated median DR_5 value based on performance metrics for all 926 polygenic risk scores and all diseases studied in the PGS Catalog. Abbreviations: PRS – polygenic risk score; CVD – cardiovascular disease; VTE – venous thromboembolic disease; AMD – age-related macular degeneration; POAG – primary open angle glaucoma; SLE – systemic lupus erythematosus.

Figure 2. Relative polygenic risk score distributions among those later affected or not by CAD (left panel) and breast cancer (right panel). The mean value of the polygenic risk score distribution those later affected is shifted 0.48 standard deviation units to the right of the mean of the distribution for those who remain unaffected in the case of CAD and 0.37 standard deviation units to the right in the case of breast cancer. The corresponding DR_5 values and odds ratios for comparisons of the top and bottom 1%. 5%, 10%, 20% and 25% of the unaffected polygenic risk score distribution are shown below each plot.

Figure 3. Likelihood ratios and 10-year odds of CAD (top panel) and breast cancer (bottom panel) for 50-year olds with a polygenic risk score result corresponding to the $2.5th$, $25th$, $75th$ and $97.5th$ centiles of the corresponding distribution.

Figure 4. Likelihood ratios, odds and the 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 CAD of 1:19 (left panel) and women with a 10-year odds of breast cancer of $1:41$ (right panel).

Figure 5. As figure 4 but comparing highest and lowest 2.5% of the unaffeted polygenic risk score distributions for CAD (left panel) and breast cancer (right panel).

Figure 6. The estimated number of breast cancer cases detected and missed, the number of false positives, and the number of additional mammograms for a two-stage screening test

medRxiv preprint doi: https://doi.org/10.1101/2022.02.18.22271049; this version posted December 19, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

using a polygenic risk score PGS000004 with a cut-off at the unaffected 97.5th centile. Estimates are based on on a hypothetical cohort of 100,000 40-year old women with a background 10-year odds of breast cancer of $1:41$. Performance of mammography in the detection of breast cancer uses estimates from JAMA. 2005; 293:1245

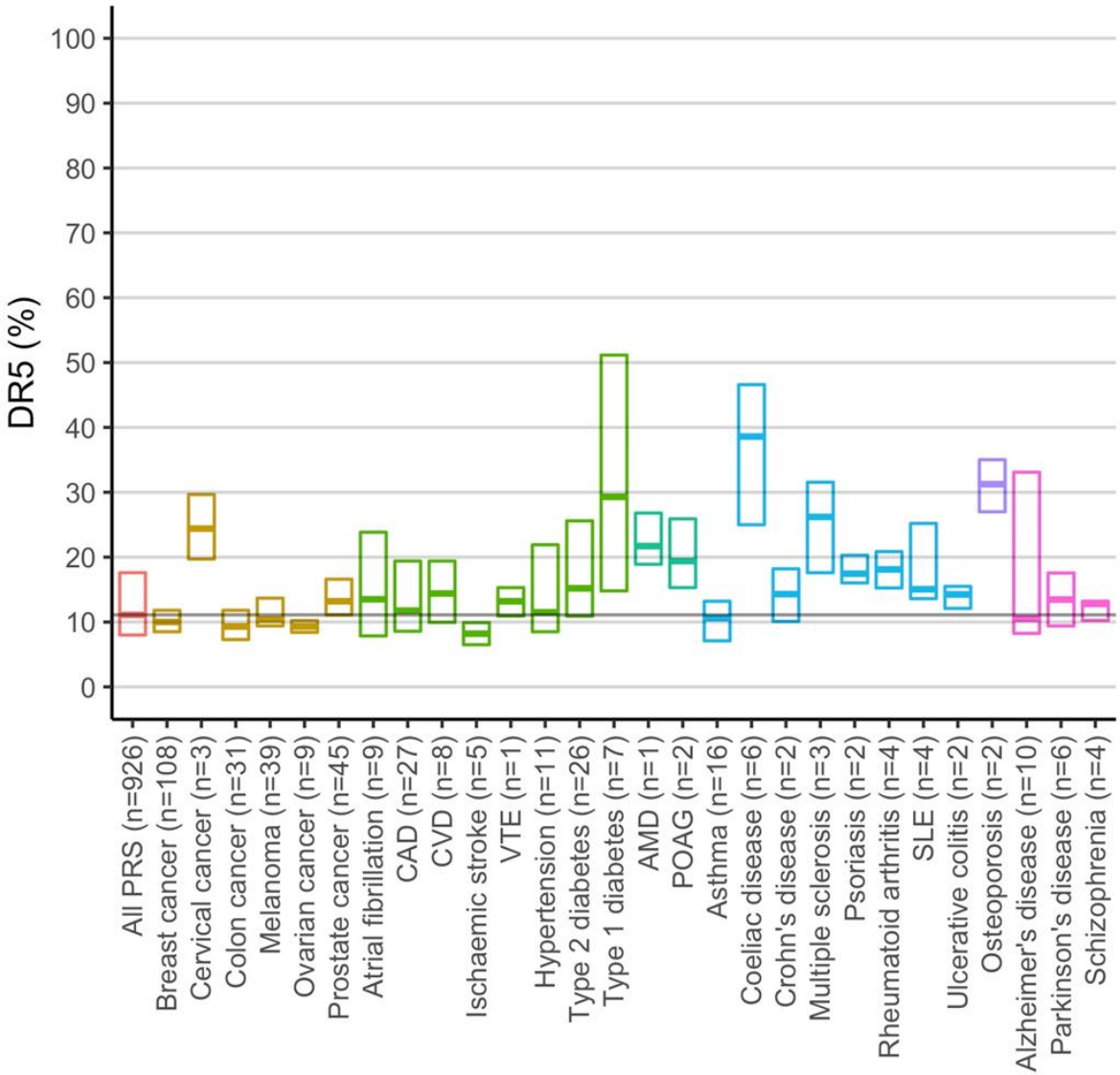
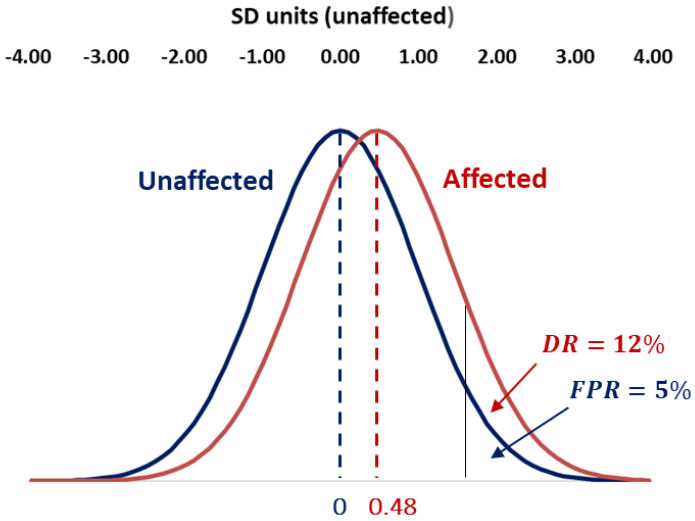
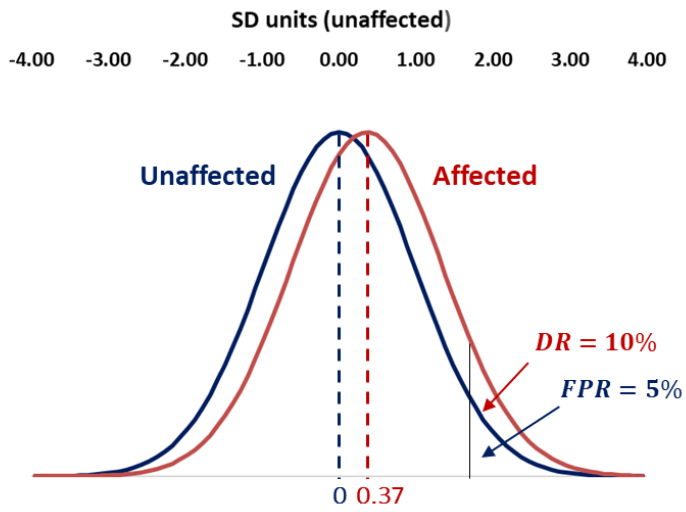

DR5 (%)
100
90
80
70
60
50
40
30
20
10
0
All PRS (n=926)
Breast cancer (n=108)
Cervical cancer (n=3)
Colon cancer (n=31)
Melanoma (n=39)
Ovarian cancer (n=9)
Prostate cancer (n=45)
Atrial fibrillation (n=9)
CAD (n=27)
CVD (n=8)
Ischaemic stroke (n=5)
VTE (n=1)
Hypertension (n=11)
Type 2 diabetes (n=26)
Type 1 diabetes (n=7)
AMD (n=1)
POAG (n=2)
Asthma (n=16)
Coeliac disease (n=6)
Crohn's disease (n=2)
Multiple sclerosis (n=3)
Psoriasis (n=2)
Rheumatoid arthritis (n=4)
SLE (n=4)
Ulcerative colitis (n=2)
Osteoporosis (n=2)
Alzheimer's disease (n=10)
Parkinson's disease (n=6)
Schizophrenia (n=4)

Figure 1



Coronary artery disease

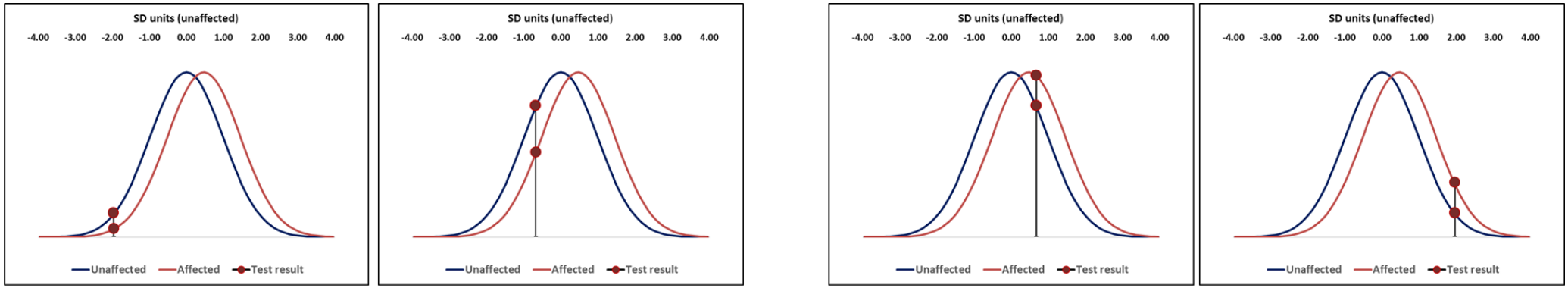
OR Top vs Bottom 1%	**13**
OR Top vs Bottom 5%	**7**
OR Top vs Bottom 10%	**5**
OR Top vs Bottom 20%	**4**
OR Top vs Bottom 25%	**3**
Detection rate	**12%**
False positive rate	**5%**



Breast cancer

OR Top vs Bottom 1%	**7**
OR Top vs Bottom 5%	**5**
OR Top vs Bottom 10%	**4**
OR Top vs Bottom 20%	**3**
OR Top vs Bottom 25%	**3**
Detection rate	**10%**
False positive rate	**5%**

Figure 2



LR 0.35
Odds 1:54

LR 0.65
Odds 1: 29

Coronary artery disease (background odds 1:19)

LR 1.23
Odds 1:15

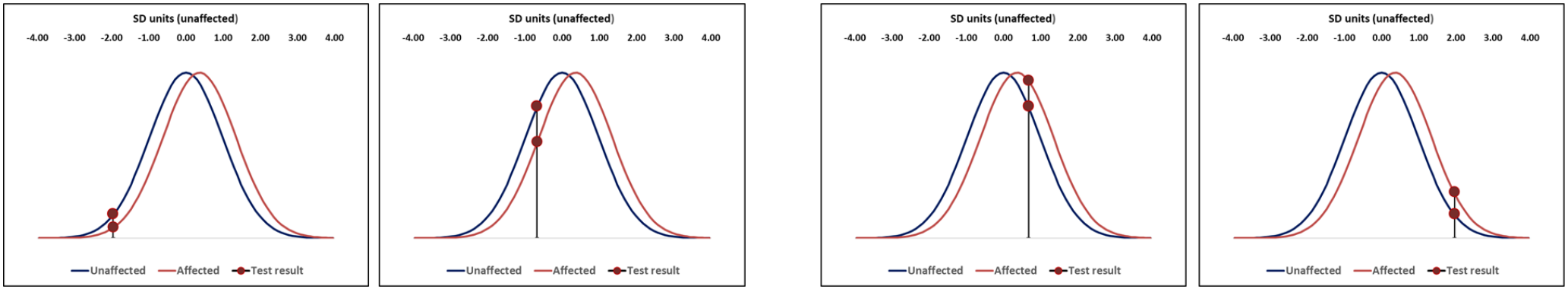
LR 2.27
Odds 1:8

2.5th

25th

75th

97.5th



LR 0.45
Odds 1:91

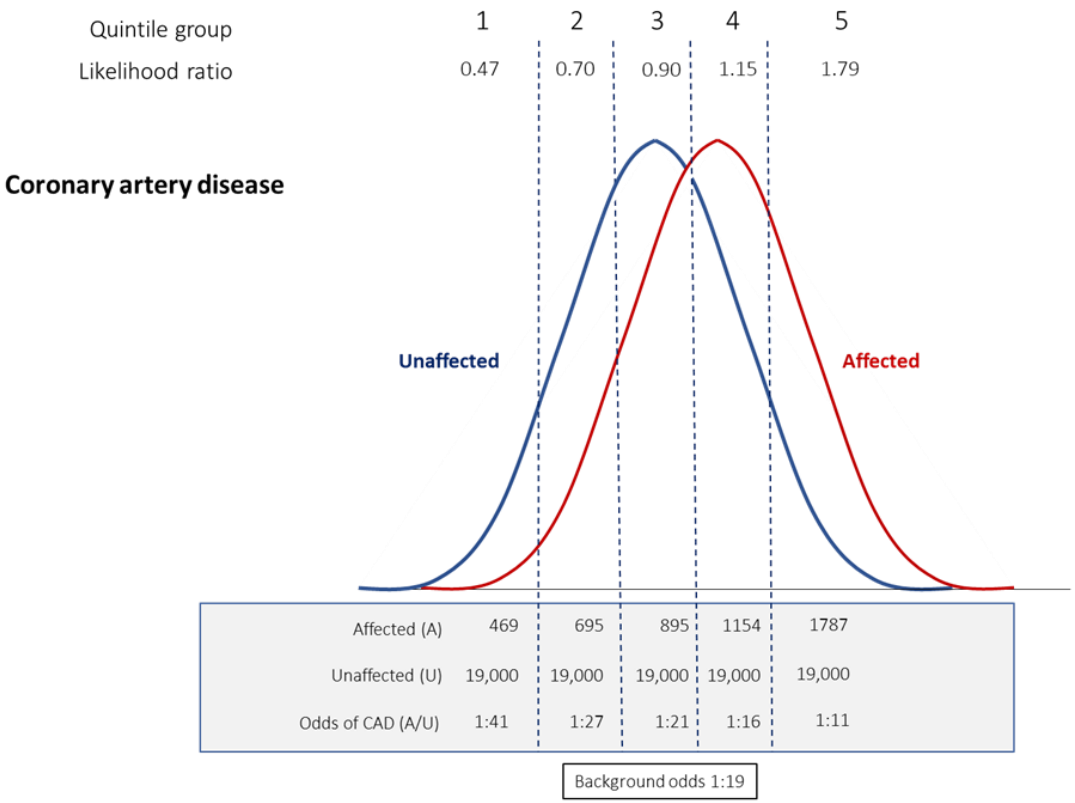
LR 0.73
Odds 1: 56

Breast cancer (background odds 1:41)

LR 1.2
Odds 1:34

LR 1.93
Odds 1:21

Figure 3


Quintile group
1
2
3
4
5
Likelihood ratio
0.47
0.70
0.90
1.15
1.79
Coronary artery disease
Unaffected
Affected
Affected (A) 469 695 895 1154 1787
Unaffected (U) 19,000 19,000 19,000 19,000 19,000
Odds of CAD (A/U) 1:41 1:27 1:21 1:16 1:11
Background odds 1:19

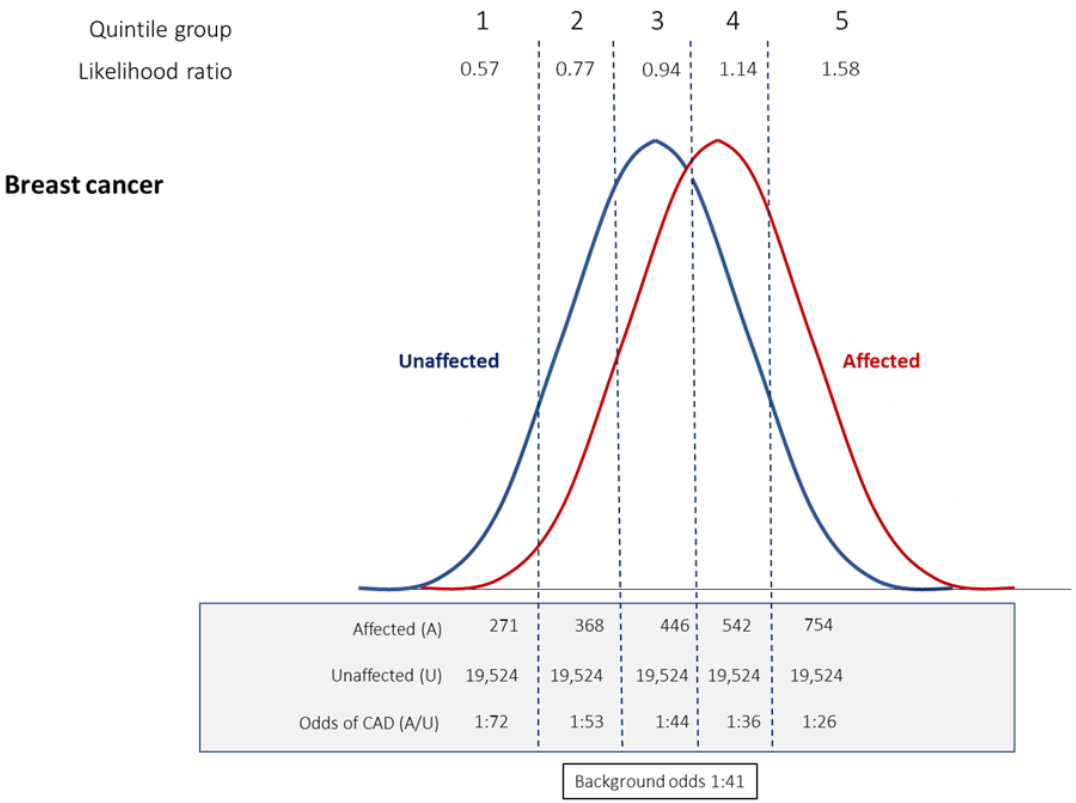
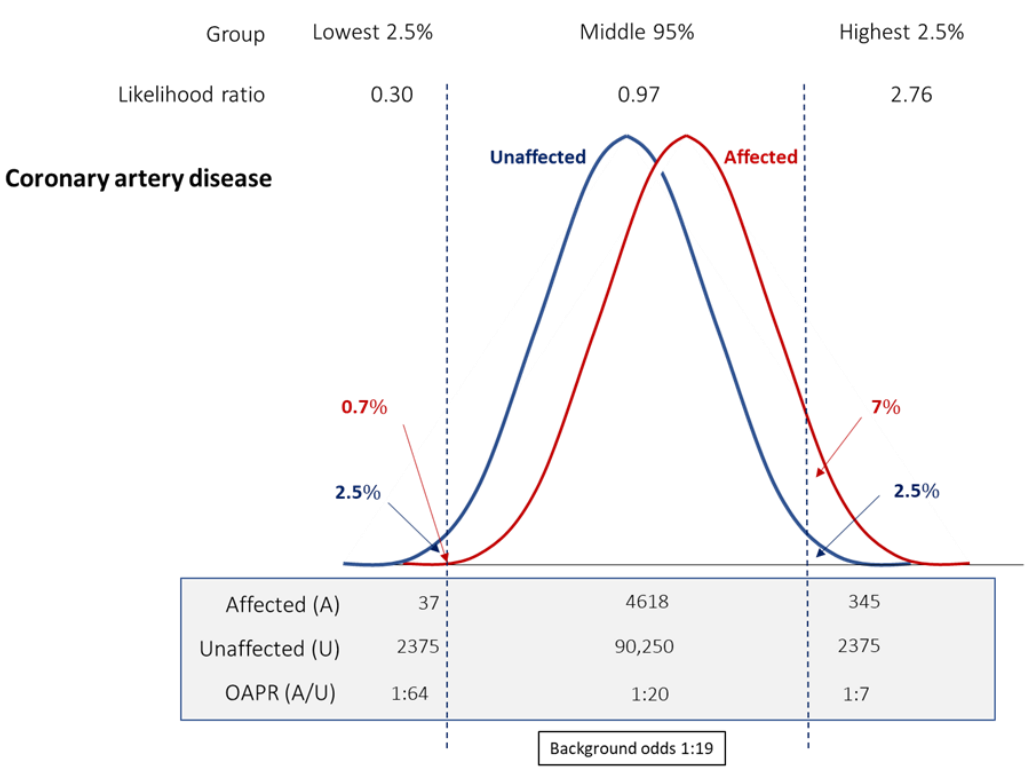

Quintile group
1
2
3
4
5
Likelihood ratio
0.57
0.77
0.94
1.14
1.58
Breast cancer
Unaffected
Affected
Affected (A) 271 368 446 542 754
Unaffected (U) 19,524 19,524 19,524 19,524 19,524
Odds of CAD (A/U) 1:72 1:53 1:44 1:36 1:26
Background odds 1:41

Figure 4


Group
Lowest 2.5%
Middle 95%
Highest 2.5%
Likelihood ratio
0.30
0.97
2.76
Coronary artery disease
Unaffected
Affected
0.7%
7%
2.5%
2.5%
Affected (A) 37 4618 345
Unaffected (U) 2375 90,250 2375
OAPR (A/U) 1:64 1:20 1:7
Background odds 1:19

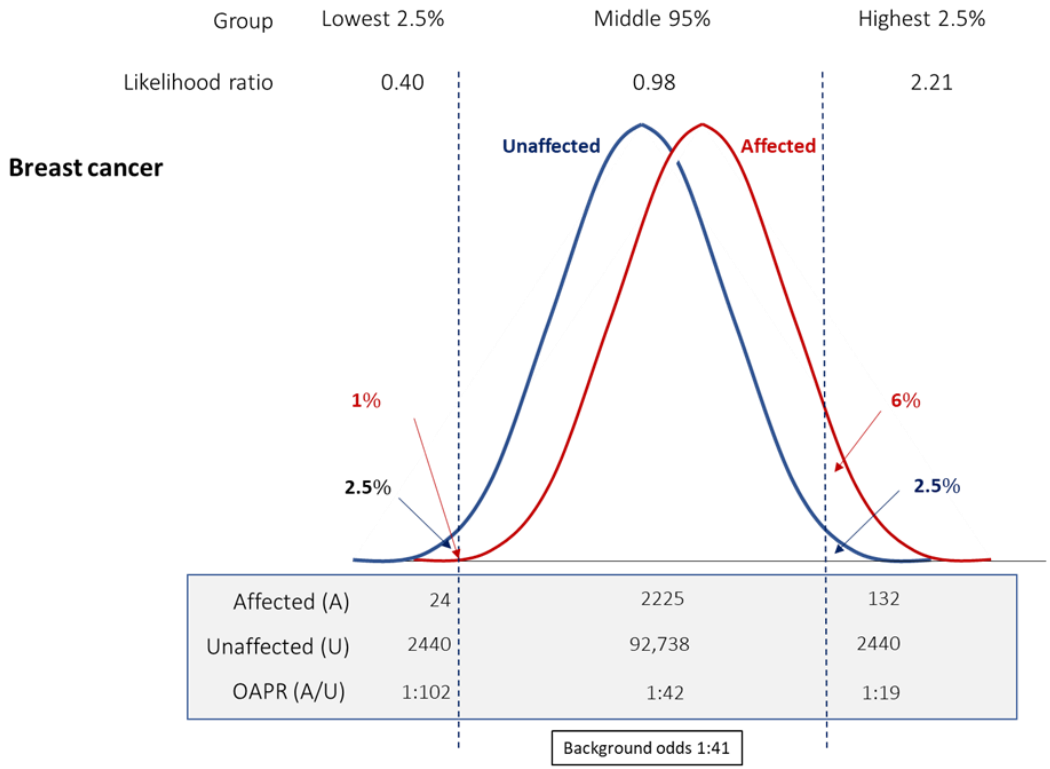

Group
Lowest 2.5%
Middle 95%
Highest 2.5%
Likelihood ratio
0.40
0.98
2.21
Breast cancer
Unaffected
Affected
1%
6%
2.5%
2.5%
Affected (A) 24 2225 132
Unaffected (U) 2440 92,738 2440
OAPR (A/U) 1:102 1:42 1:19
Background odds 1:41

Figure 5

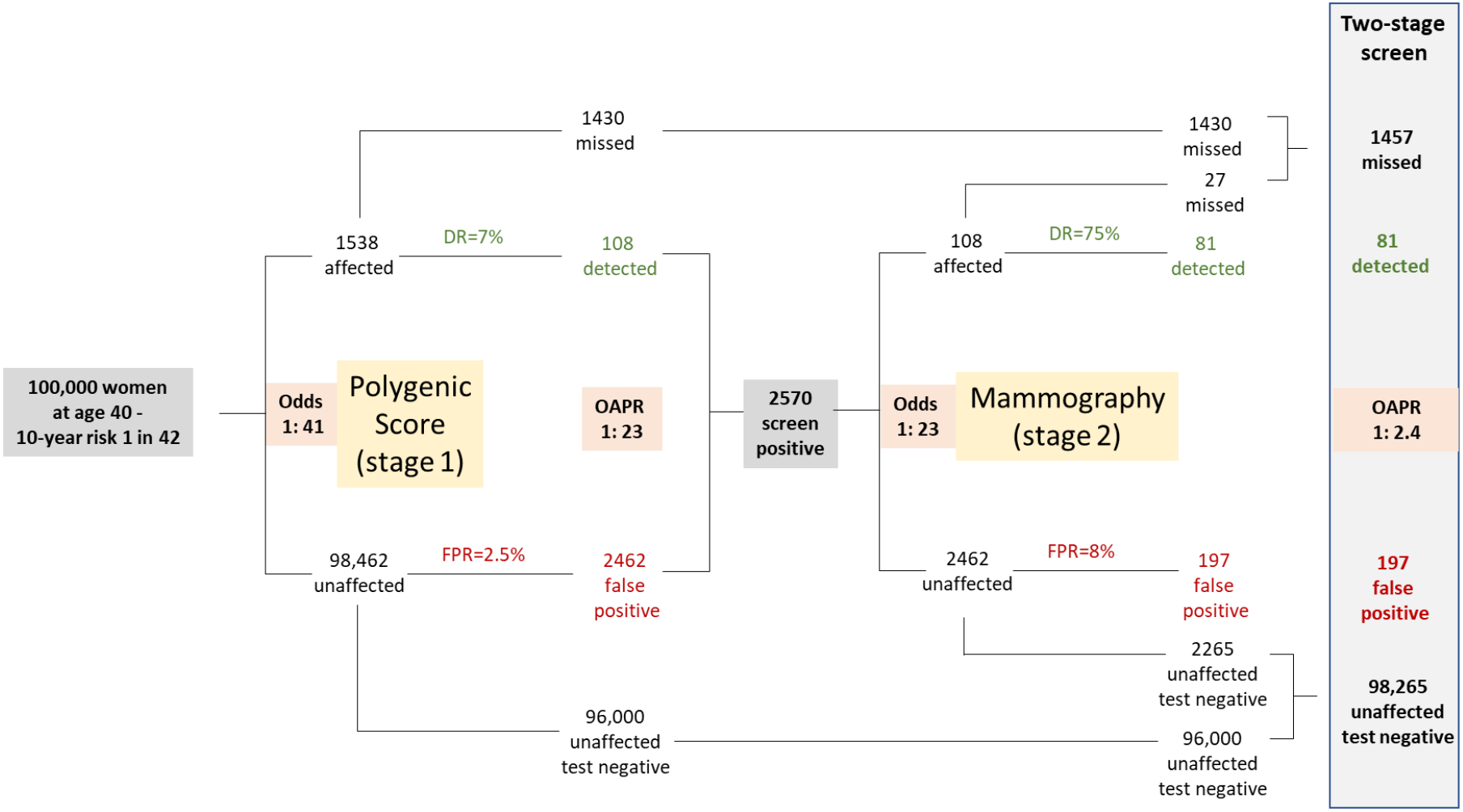

Two-stage screen
100,000 women at age 40 - 10-year risk 1 in 42
Odds 1: 41
Polygenic Score (stage 1)
1538 affected
DR=7%
108 detected
1430 missed
OAPR 1: 23
98,462 unaffected
FPR=2.5%
2462 false positive
96,000 unaffected test negative
2570 screen positive
Odds 1: 23
Mammography (stage 2)
108 affected
DR=75%
81 detected
27 missed
2462 unaffected
FPR=8%
197 false positive
2265 unaffected test negative
96,000 unaffected test negative
1430 missed
1457 missed
81 detected
OAPR 1: 2.4
197 false positive
98,265 unaffected test negative

Figure 6