**Marginal role for 53 common genetic variants in cardiovascular disease prediction**

Richard W Morris (1,2), Jackie A Cooper (3), Tina Shah (4), Andrew Wong (5), Fotios Drenos (4,6), Jorgen Engmann (4), Stela McLachlan (7), Barbara Jefferis (2), Caroline Dale (8), Rebecca Hardy (5), Diana Kuh (5), Yoav Ben-Shlomo (1), S Goya Wannamethee (2), Peter H Whincup (9), Juan-Pablo Casas (3), Mika Kivimaki (10), Meena Kumari (10, 11), Philippa J Talmud (3), Jacqueline F Price (7), Frank Dudbridge (8), Aroon D Hingorani (4), Steve E Humphries (3)

on behalf of the UCLEB consortium

Corresponding author:

Professor RW Morris, School of Social & Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK.

Tel: +44 117 331 3935

Email: [richard.morris@bristol.ac.uk](mailto:richard.morris@bristol.ac.uk)

Institutional affiliations:

1 School of Social & Community Medicine, University of Bristol, Bristol, UK

2 Department of Primary Care & Population Health, University College London, UK

3 Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, UK

4 Institute of Cardiovascular Science and Farr Institute, University College London, London, United Kingdom

5 MRC Unit for Lifelong Health and Ageing at UCL, London, United Kingdom

6 MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

7 Centre for Population Health Sciences, University of Edinburgh, Edinburgh, United Kingdom

8 Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, UK

9 Division of Population Health Sciences and Education, St George’s, University of London, London, United Kingdom

10 Department of Epidemiology & Public Health, UCL Institute of Epidemiology & Health Care, University College London, London, United Kingdom

11 Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom

**ABSTRACT**

**Objective**: We investigated discrimination and calibration of cardiovascular disease (CVD) risk scores when genotypic was added to phenotypic information. The potential of genetic information for those at intermediate risk by a phenotype-based risk score was assessed.

**Methods:** Data were from seven prospective studies including 11,851 individuals initially free of CVD or diabetes, with 1,444 incident CVD events over 10 years’ follow-up. We calculated a score from 53 CVD-related single nucleotide polymorphisms (SNPs) and an established CVD risk equation “QRISK-2” comprising phenotypic measures. The area under the receiver operating characteristic curve (AUROC), detection rate for given false positive rate (FPR), and net reclassification index (NRI) were estimated for gene scores alone and in addition to the QRISK-2 CVD risk score. We also evaluated use of genetic information only for those at intermediate risk according to QRISK-2.

**Results**: The AUROC was 0.635 for QRISK-2 alone, and 0.623 with addition of the gene score. The detection rate for 5% FPR improved from 11.9% to 12.0% when adding the gene score. For a 10-year CVD risk cut-off point of 10%, the NRI was 0.25% when the gene score was added to QRISK-2. Applying the genetic risk score only to those with QRISK-2 risk of 10-<20%, and prescribing statins where risk exceeded 20%, suggested genetic information could prevent one additional event for every 462 people screened.

**Conclusions:** The gene score produced minimal incremental population-wide utility over phenotypic risk prediction of CVD. Tailored prediction using genetic information for those at intermediate risk may have clinical utility.

*(250 words)*

**KEY MESSAGES**

**What is already known about this subject?**

Predictive accuracy of cardiovascular risk, generally based on well-established phenotypic measures, has often seemed disappointing. Genome-wide association studies have highlighted new genetic loci related to coronary artery disease and stroke.

**What does this study add?**

When information on 53 single nucleotide polymorphisms about individuals from seven UK prospective studies are added to a well-established cardiovascular risk score, the ability to predict cardiovascular disease (CVD) over the next 10 years is not enhanced.

However if a genetic risk score is applied to individuals classed at intermediate risk according to a traditional risk score, some individuals will be re-classified at high risk and CVD events will be postponed due to timely use of lipid-lowering therapy. This two-stage strategy will postpone 216 events in every 100,000 people screened.

**How might this impact on clinical practice?**

Routine use of genetic profiles is not necessary for everyone screened for cardiovascular risk. However there may be clinical utility for a genetic risk score for those initially screened as of intermediate risk.

**INTRODUCTION**

Despite the importance of predicting future cardiovascular disease (CVD) among initially healthy adults, predictive accuracy has often seemed disappointing, as most individuals who eventually suffer a CVD event were previously at average rather than high risk: the prevention paradox**1**. Lowering cholesterol through statin use reduces CVD risk2. Accordingly, several major guidelines3-6 recommend lipid lowering therapy for people with a raised 10-year CVD predicted risk, traditionally using a threshold of 20%. However, with recent patent expiries resulting in reduced acquisition cost, and increasing evidence on the limited harms of statins, the 10-year CVD risk threshold for primary prevention of CVD has been reduced to 10% in the UK3 and to 7.5% in the USA4. However, these decisions have been questioned, especially since people with intermediate 10-year CVD risk (e.g. 10-20%) may be reluctant to undergo statin therapy5. Refining risk estimation may be of particular interest in such individuals, as well as helping guide appropriate targeting of alternative therapies currently under development.

Considerable advances have taken place in understanding genetic determinants of CVD in recent years, and the CardiogramPlusC4D collaboration have now catalogued associations of hundreds of thousands of single nucleotide polymorphisms (SNPs) across the genome , using data on over 63,000 coronary heart disease (CHD) cases and 130,000 controls6. This collaboration identified 46 loci containing SNPs that surpassed genome-wide levels of statistical significance. Further SNPs associated with ischaemic stroke risk have included rs783396 from the AIM1 gene in chromosome 6q217, and rs12425791 (closest gene NINJ2, chromosome 12)8. Case-control studies do not permit estimation of absolute risk. We therefore evaluated the predictive performance of a gene score based on 53 SNPs associated with CHD or stroke on its own and in conjunction with the established non genetic QRISK-2 risk tool9 (developed for CVD prediction in UK populations), using the University College-London School-Edinburgh-Bristol (UCLEB) Consortium of prospective population studies10.

**METHODS**

**University College London, London School of Hygiene and Tropical Medicine, Edinburgh and Bristol**

**(UCLEB) Consortium**

A full description of the UCLEB Consortium has been previously published10. Briefly, the studies comprise individuals almost exclusively of European ancestry from a wide geographic range within the UK. For the current analysis, seven prospective studies with genotype and complete information on CVD incidence were included. For full details of individual studies, see Online Supplementary information. In four of the studies (EAS, NSHD, WHII and CaPS), all participants providing blood samples were genotyped, but a nested case-control sample was used for the remainder. Analysis was restricted to 11,851 individuals aged 85 years or less and excluded 1,542 individuals with prevalent diabetes and 1,191 with prevalent CVD.

Informed consent was obtained for all subjects included in UCLEB research. Written approval from individual Research Ethics Committees to use anonymised individual level data has been obtained by each participating study.

**Clinical characteristics of the participants**

Within individual cohorts, biochemical measurements were performed in accredited laboratories using international standards. 10. For the current analysis, earliest available measurements were abstracted for each study on relevant phenotypes. Medication data included lipid lowering drugs (statins or other), and blood pressure lowering drugs: for the latter, adjustment was made by adding 15 mmHg for systolic, and 10 mmHg for diastolic blood pressure11.

**Definition of cardiovascular disease**

The definition of prevalent CVD (from the same time point as the phenotypic measurements) was based on either self-report, medical record review or examination with ECG. CVD consisted of a combination of coronary heart disease (CHD) and stroke. CHD included all non-fatal myocardial infarction (MI) or any revascularisation procedure (coronary artery bypass surgery or angioplasty) and fatal CHD. Stroke included all nonfatal stroke (ischaemic & haemorrhagic combined, but excluding transient ischaemic attacks) and fatal stroke. Fatal events were classed according to ICD-10 codes: I20-I25 for CHD, I60-I69 for stroke.

**Genotyping**

DNA was extracted from blood samples either collected at baseline (BWHHS) or at a subsequent resurvey (BRHS, MRC NSHD, EAS, WHII, ELSA, CaPS)10. Genotype data were based on the Illumina CardioMetabochip which incorporates approximately 200,000 SNPs from loci previously identified for associations with cardiometabolic disease risk factors and outcomes12. Imputation was conducted against the 1000 genomes reference panel, providing information on approximately 2 million typed or imputed SNPs. Duplicate samples were genotyped to compute the error rate. Quality control on genotyped samples has been previously reported10, and all included SNPs had a call rate of >98%. Genotypes were in Hardy Weinberg Equilibrium in all studies.

We used the list of CVD risk SNPs recently identified in large meta-analyses of CHD 6 and stroke7 8, (Supplementary File, eTable 1);all 53 CVD SNPs except one were typed through the CardioMetabochip: one SNP associated with stroke (rs783396) was imputed.

**Statistical analysis**

*Score construction*

We used the QRISK-2 2014 batch processor using data for age, sex, smoking, family history of CVD, body mass index (BMI), blood pressure, treatment for hypertension, total and HDL-cholesterol, to compute the QRISK-2 risk probabilities9. We computed a genetic risk score (GRS) weighted according to published coefficients (log odds ratios) for the 53 SNPs 6. Coefficients were multiplied by 0, 1, or 2 according to the number of risk alleles carried by each person. The logits of the QRISK-2 probabilities were added to the GRS to produce a combined score. As a sensitivity analysis, to address concerns that beta coefficients for the individual SNPs selected for the GRS may be inflated, we calculated an unweighted gene score, and followed similar procedures.

*Association testing*

Logistic regression models were fitted to obtain the odds ratio (OR) per standard deviation (SD) increase in the GRS as well as OR associated with each quintile. Association models were fitted using the combined dataset with a term for study included as a fixed effect.

*Model Discrimination*

We calculated the area under the receiver operator characteristic curve (AUROC), and the detection rate, defined as the proportion of all cases detected for a false positive rate (FPR) of 5% (DR5) and 10% (DR10). AUROCs were calculated separately for each study and combined using both fixed effects and random effects meta-analysis. Improvements in discrimination were assessed by calculating the difference between the two AUROCs in each study with bootstrap estimates of the confidence interval and then combining these over the studies.

*Model calibration*

For the combined score estimates of risk were obtained by converting the logit back to a probability. For all studies but ELSA, the number of events occurring within 10 years of baseline was observed. For ELSA, since follow up was for 5 years only, we doubled this to give the 10 year observed risk. Observed risks were then compared to predicted risks within tenths of the predicted risk distribution, and the Hosmer-Lemeshow test was used to assess goodness of fit.

*Reclassification of CVD risk*

We used the net reclassification improvement (NRI) to evaluate improvement in risk prediction. This metric quantifies the extent to which the combined score moved people to risk categories that better reflected their future event status13. In three of the studies, all cases were genotyped but only a fraction of the controls so it was necessary to up-weight data for controls to reflect properly the proportion of cases in the population. For example, if within a particular age group of one study, only 80% of controls had been selected for genotyping, we assigned a weight of 1.25 (=100/80) to all those controls but a weight of 1 to cases, when calculating the number who had been reclassified. We used three 10-year CVD risk categories (<10%, 10%-19.9%, and 20% or higher). We calculated the NRI without accounting for study, and then calculated NRI and its standard error for each study and combined into an overall NRI with a fixed effects meta-analysis. There was very little difference in the two methods so we present results for the latter.

We also followed the Emerging Risk Factors Collaboration’s method14 in assessing additional predictive value of novel risk factors for individuals initially categorised as intermediate risk according to established risk factors. Of those whose predicted risk was between 10 and 20% according to the QRISK-2 equation, we calculated the number who would subsequently be reclassified as high risk once the GRS was added. We assumed all such individuals would be treated with statins and would achieve a 20% relative risk reduction (adherence assumed to be similar to that seen in trials2), and from this we estimated the absolute number of cardiovascular events that might be prevented. This enabled us to calculate the number needed to screen to prevent one event.

All analysis was conducted using Stata version 13.1 (StataCorp, Texas).

**RESULTS**

*Characteristics of the study participants*

Studies differed by sex and age (Table 1). A total of 1,444 individuals out of 11,851 (1054 CHD events, 390 strokes) experienced CVD within 10 years of follow up (Figure 1).  297 events were fatal. The 10-year CVD event rates varied by study, from 4.7% in NSHD (mean age 53 years at baseline of follow-up) to 37.2% in EAS (mean age 64.2 years). Only 165 (1.4%) of the participants were on statin treatment at the start of follow up.

*Genetic risk score and association with CVD risk factors and CVD events*

Not every SNP demonstrated similar associations with CVD in the UCLEB data to those previously published (Supplementary File, eTable 1), with ORs <1 for 14 of the 53 SNPs in the UCLEB data.

There was a clear positive relationship of the GRS with total cholesterol and an inverse relationship with HDL cholesterol (Supplementary File, eTable2). These associations attenuated when eight SNPs related to LDL concentration were excluded from the gene score. Only a very modest positive association was seen with reported family history.

Odds ratios of incident CVD for successive quintiles of the GRS compared with the lowest quintile were 0.88, 1.10, 1.12 and 1.15 respectively, with an OR of 1.09 per SD increase (95%CI 1.03-1.15, p=0.005). Restricting incident CVD cases to 137 fatal events within 10 years, the OR for the GRS per SD increase was 1.03 (95%CI 0.87-1.22, p=0.74). When considering prevalent CVD cases, the equivalent OR was 1.17 (95%CI 1.10-1.25, p=8.2x10-7). The relationship of the QRISK-2 score with all incident CVD events was much stronger (OR per SD increase 1.92: 95%CI 1.78-2.08, p=2.6x10-58).

*Predictive accuracy of the genetic risk score alone and in combination with QRISK-2*

Table 2 shows the AUROCs, for the GRS (0.524) and QRISK (0.635) alone and the two in combination (0.623; see also eTable 3), as well as the detection rates for 5% and 10% false positive rates (FPRs). These AUROC estimates were virtually identical when family history data were not used for the QRISK-2 score, and also when random effects, instead of fixed effects analysis, was used to combine studies’ results. Detection rates for 5% and 10% FPRs were 6.8% and 13.1% respectively for the GRS alone. The corresponding detection rates for QRISK2 were 11.9% and 21.2%, changing to 12.0% and 19.6% respectively when the GRS was added.

Figure 2 shows that although QRISK-2 was well calibrated with observed risk over the majority of the risk distribution, it modestly under-predicted at low levels of risk and substantially over-predicted risk for those in the top three tenths of the predicted risk distribution. Adding information from the GRS had little effect on calibration: both predictive scores departed significantly from being well calibrated (28=309.0 for QRISK-2 and 427.1 for QRISK-2 + GRS by the Hosmer-Lemeshow test).

*Reclassification*

Net reclassification indices are shown in Table 3 according to whether individuals were above or below 10% predicted risk (part A), and whether above or below 20% predicted risk (part B). For those who did not actually experience an event, an extra 2.33% individuals crossed the threshold downwards rather than upwards when the GRS was added to the QRISK-2 equation. For those who did experience an event, an extra 2.08% individuals crossed the threshold downwards rather than upwards when the GRS was added. Overall, the NRI was therefore 0.25% (95%CI -1.33 to 1.83%). When a threshold of 20% was used, a net increase of 0.90% was observed for those who crossed the threshold downwards rather than upwards among those who did not experience events, but also a net increase of 0.25% in the same direction for those who did experience events. Hence the NRI was 0.65%.

*Estimated performance of a sequential screening strategy*

Figure 3 illustrates the estimated effect of a sequential screening strategy applied to 100,000 people screened for CVD risk using QRISK-2 followed by addition of information from a GRS among those estimated to be in the intermediate risk category (10 year risk 10% to <20%). Based on QRISK-2 estimates, for every 100,000 people in the population from which our data were drawn, 29,445 would be at intermediate risk. When adding the GRS, 16,782 would remain as intermediate risk; 7,229 would be reclassified as low risk, and 5,434 would be reclassified as high risk, thus making them eligible for statin treatment. Based on extrapolation from the current analysis of those reclassified by addition of the GRS, an estimated 1,082 would go on to suffer a CVD event within 10 years. Assuming a 20% reduction in events from statin treatment, 216 events (20% of 1082) would be expected to be prevented. Therefore, adding information from the GRS to QRISK-2 among those classified as being at intermediate risk by the latter would postpone one event for every 462 screened.

*Potential influence of age on screening performance*

Discrimination and reclassification was estimated separately for participants aged under, and over age 60 (Supplementary File, eTables 4-6). There was no evidence of any differences in AUROC for the GRS alone (0.530 and 0.518 respectively), in the improvement in AUROC of GRS compared with QRISK-2 alone (-0.010 and -0.007 respectively), NRI based on the 10% cut off point for predicted risk (0.60% for each age group), or NRI based on the 20% cut point (1.0% and 1.5%).

**DISCUSSION**

Our study suggests that gene scores from 53 SNPs were not effective in predicting 10 year risk of CVD, with an AUC of only 0.524; this area was 0.635 for a model with QRISK-2 alone and 0.623 when a genetic risk score was added into the model. Nevertheless, the GRS appeared to carry some utility when applied only to those who, according to conventional risk scoring, would have been classified at intermediate risk, by moving some individuals into the high-risk category. Among 100,000 people from a population represented by our combined studies, 29,445 would be classed as of intermediate risk according to the QRISK equation, but of these, 5,434 would then be reclassified as high risk once the GRS was applied, and 1,082 would suffer a CVD event if untreated with statins. For 20% of these, 216, treatment with statins according to guidelines could postpone the CVD event, leaving a number needed to screen of 462 to postpone one CVD event. Recent evidence suggests the risk reductions from statin therapy might be enhanced for those at highest risk15, so this figure may be conservative.

Our data include seven British prospective studies, in which CVD events were defined in a standard manner10, genotyping followed a common protocol and almost 1,500 incident CVD events were available for analysis. The participants of the studies were a median of 53 years and more commonly male. This age group represents a population group most eligible for cardiovascular screening, and we did not see differential performance of the screening algorithms according to age group, even when we restricted analysis to those aged 53 or less (data not shown). Genetic information may be more useful for those aged under 40 (not represented in this study), but a lifetime risk equation would then be required. In all seven studies combined, we noticed substantial over-prediction by QRISK-2, despite its apparently good calibration in other UK based prospective studies16. Thus, while we noted a two-stage screening procedure would yield identification and treatment of some high risk individuals who would have been classified at intermediate risk by QRISK-2, the phenomenon of over-prediction by QRISK-2 suggests that many more needed to be reclassified as low risk. The genetic score did not actually improve the calibration at all.

In constructing the GRS, we used regression coefficients catalogued by the CardiogramplusC4D consortium on a very large dataset. While the regression coefficients for SNPs extracted from this dataset will perform less optimally when applied to a new dataset, we believe this represents a truer test of validation17. The 53 SNPs will probably be those SNPs most strongly associated with CVD that will ever be found, but their combined effect still represents a small proportion of heritability of CVD and is still small compared with major phenotypic risk factors. Better prediction from genotypic information may be expected from identification of several thousand more SNPs18.

The development of QRISK-2, and most of our studies’ baselines, pre-dated the statin era and the proportion taking statins during follow-up would be modest. Our data are able to evaluate what risks could have been prevented had statins been widely available.

Other attempts to evaluate use of genotypic data for cardiovascular risk screening have been made. A marginal improvement in discrimination over and above the predictive power of traditional coronary risk factors was found in the ARIC study for black participants (but less clearly for white participants)19, and among European men20. Among participants of the Framingham study, no significant improvement in discrimination was found but a modest benefit in reclassification of CVD risk21. The Framingham study and the REGICOR study (north eastern Spain, low CHD risk) were used to assess CHD risk: this showed that a GRS improved discrimination for Framingham participants but not REGICOR22. However, better performance was seen for reclassification of those at intermediate risk in both studies. The same was true in the FINRISK studies23, who estimated with a two-stage screening that 135 events could be prevented among 100,000 screened, slightly less than 216/100,000 in the present study.

Recent data from the Malmὂ Diet and Cancer Study showed that family history did not lessen the predictive utility of a GRS, but the GRS added predictive value over phenotypic risk scores which included family history24. By contrast, our data find little evidence for improvement in discrimination over a phenotypic risk score, whether or not it includes family history.

The Rotterdam study25 conducted similar GRS analysis using the same subset of 53 as in the present study. As in our study, a stronger relationship of gene score was observed for prevalent cases than incident. The present study also observed a weaker relationship of gene score with CVD mortality; therefore supporting the suggestion that some genes identified by CardiogramplusC4D were related to better survival after CVD, rather than to incident disease, and questions the generation of signals through genome-wide association studies in case-control studies, if no distinction can be made between cases who have died and those who survived. A fully powered prospective study is required of individuals with incident CVD, to compare genotypes between survivors and those who died of the event.

Our findings underline the relatively disappointing performance of gene scores in adding to cardiovascular risk scores based on established risk factors. Nevertheless, we have shown the potential for refining risk calculation in those initially classed as of intermediate risk. A similar analysis applied to selective use of CRP and fibrinogen for those at intermediate risk suggested these markers would require over 3,000 screened to postpone a CVD event14: the relatively better performance of the GRS in the present study is because a higher proportion of those at intermediate risk were reclassified as high risk. It has been shown that a collection of alternative risk scores (including QRISK-2), based on established risk factors, are liable to disagree over classifying individuals as high risk26. Therefore including a GRS may help identify an intermediate group who should properly be classed at high risk. Despite current UK recommendations that treatment with statins be extended to those at intermediate risk3 (CVD risk 10-20% over 10 years) as well as those at high risk (over 20%), family physicians may be reluctant to do so. The Joint British Societies’ (JBS3) consensus recommendations for the prevention of CVD did not recommend the use of genetic information, which was seen as currently performing less well than established risk factors27. However for individuals not meeting the criteria for lifestyle or drug therapy, JBS3 recommended calculation of metrics such as heart age, relating to lifetime risk. A gene score with good predictive power would seem particularly suitable to evaluate lifetime risk, given its non-modifiable nature throughout the life course.

The Rotterdam study25 constructed a second risk score based on 169 SNPs including the original 53 modelled in our study, as well as a further 116 for whom only modestly significant changes in risks were demonstrated. This second risk score performed better than the first, and further gene discovery may therefore produce greater improvements. However at present our results and those of others cannot support the population-wide use of genetic risk scores in targeting treatment, despite the modest utility in reclassifying those at intermediate risk.

**Funding:**

The UCLEB Consortium is supported by a British Heart Foundation Programme Grant (RG/10/12/28456). BRHS is a British Heart Foundation Research Group and is supported by British Heart Foundation (RG/08/013/25942). The WHII study is supported by grants from the Medical Research Council (K013351; ID85374), British Heart Foundation (RG/07/008/23674), Stroke Association, National Heart Lung and Blood Institute (HL036310), National Institute on Aging (5RO1AG13196) Agency for Health Care Policy Research (HS06516), and the John D. and Catherine T. MacArthur Foundation Research Networks on Successful Midlife Development and Socioeconomic Status and Health. Samples from the ELSA DNA Repository (EDNAR), received support under a grant (AG1764406S1) awarded by the National Institute on Ageing (NIA). ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute of Fiscal Studies. The data were collected by the National Centre for Social Research. MRC NSHD is funded by the Medical Research Council [MC\_UU\_12019/1]. BWHHS is supported by funding from the British Heart Foundation and the Department of Health Policy Research Programme (England). EAS is funded by the British Heart Foundation (Programme Grant RG/98002), with Metabochip genotyping funded by a project grant from the Chief Scientist Office of Scotland (Project Grant CZB/4/672). CaPS was funded by the Medical Research Council and undertaken by the former MRC Epidemiology Unit (South Wales). The DNA bank was established with funding from a MRC project grant. The data archive is maintained by the University of Bristol. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. SEH holds a chair funded by the British Heart Foundation. PJT, MKu**,** ADH, and SEH were supported by the BHF (grant numbers PG07/133/24260, BHFPG08/008).

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in HEART editions and any other BMJPGL products to exploit all subsidiary rights.

**REFERENCES**

1. Rose G. Strategy of prevention: lessons from cardiovascular disease. British medical journal 1981;**282**(6279):1847-51.

2. Baigent C, Keech A, Kearney PM, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 2005;**366**(9493):1267-78.

3. National Institute for Health Care and Excellence.. Lipid Modification: Cardiovascular Risk Assessment and the Modification of Blood Lipids for the Primary and Secondary Prevention of Cardiovascular Disease. London, 2014.

4. Goff DC, Jr., Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014;**129**(25 Suppl 2):S49-73.

5. Usher-Smith JA, Pritchard J, Poole S, et al. Offering statins to a population attending health checks with a 10-year cardiovascular disease risk between 10% and 20. International journal of clinical practice 2015.

6. CardiogramplusC4D Consortium, Deloukas P, Kanoni S, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nature genetics 2013;**45**(1):25-33.

7. Matarin M, Brown WM, Scholz S, et al. A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. The Lancet Neurology 2007;**6**(5):414-20.

8. Ikram MA, Seshadri S, Bis JC, et al. Genomewide association studies of stroke. The New England journal of medicine 2009;**360**(17):1718-28.

9. Hippisley-Cox J, Coupland C, Robson J, et al. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;**341**:c6624.

10. Shah T, Engmann J, Dale C, et al. Population genomics of cardiometabolic traits: design of the University College London-London School of Hygiene and Tropical Medicine-Edinburgh-Bristol (UCLEB) Consortium. PLoS One 2013;**8**(8):e71345.

11. Tobin MD, Sheehan NA, Scurrah KJ, et al. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Statistics in medicine 2005;**24**(19):2911-35.

12. Voight BF, Kang HM, Ding J, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS genetics 2012;**8**(8):e1002793.

13. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine 2008;**27**(2):157-72; discussion 207-12.

14. Emerging Risk Factors Collaboration, Kaptoge S, Di Angelantonio E, et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction. The New England journal of medicine 2012;**367**(14):1310-20.

15. Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet 2015;**385**(9984):2264-71.

16. Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 2012;**344**:e4181.

17. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;**98**(9):691-8.

18. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS genetics 2013;**9**(3):e1003348.

19. Morrison AC, Bare LA, Chambless LE, et al. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. Am J Epidemiol 2007;**166**(1):28-35.

20. Hughes MF, Saarela O, Stritzke J, et al. Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. PLoS One 2012;**7**(7):e40922.

21. Thanassoulis G, Peloso GM, Pencina MJ, et al. A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. Circulation Cardiovascular genetics 2012;**5**(1):113-21.

22. Lluis-Ganella C, Subirana I, Lucas G, et al. Assessment of the value of a genetic risk score in improving the estimation of coronary risk. Atherosclerosis 2012;**222**(2):456-63.

23. Tikkanen E, Havulinna AS, Palotie A, et al. Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. Arteriosclerosis, thrombosis, and vascular biology 2013;**33**(9):2261-6.

24. Tada H, Melander O, Louie JZ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. European heart journal 2016;**37**(6):561-7.

25. de Vries PS, Kavousi M, Ligthart S, et al. Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: the Rotterdam Study. Int J Epidemiol 2015;**44**(2):682-8.

26. van Staa TP, Gulliford M, Ng ES, et al. Prediction of cardiovascular risk using Framingham, ASSIGN and QRISK2: how well do they predict individual rather than population risk? PLoS One 2014;**9**(10):e106455.

27. JBS Board. Joint British Societies' consensus recommendations for the prevention of cardiovascular disease (JBS3). Heart 2014;**100 Suppl 2**:ii1-ii67.

**Table 1: Characteristics of participants in the seven studies. Mean (SD) tabulated for continuous variables, % for binary variables**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | BRHS | BWHHS | CAPS | EAS | ELSA | NSHD | WHII | Total |
| Total | **2138** | **1631** | **1121** | **632** | **1184** | **2330** | **2815** | **11851** |
| Follow-up (years) | **10** | **10** | **10** | **10** | **~5** | **10** | **10** |  |
| No. of CVD events during follow up | **205** | **268** | **119** | **235** | **142** | **109** | **366** | **1444** |
| 10 year CVD event rate  (per 100 person years) | **8.4\*** | **11.4\*** | **14.7** | **37.2** | **20.8\*+** | **4.7** | **15.2** | **13.6** |
| Predicted 10-year CVD risk (QRISK-2)  (per 100 person years) | **8.4** | **26.7** | **18.1** | **20.6** | **30.2** | **10.7** | **8.2** | **15.1** |
| Age (years) | **48.9 (5.6)** | **70.7 (5.3)** | **56.7 (4.4)** | **64.2 (5.7)** | **71.5 (8.5)** | **53.0 (0.0)** | **48.8 (6.0)** |  |
| Sex, % male | **100** | **0** | **100** | **46.4** | **51.9** | **49.4** | **77.2** |  |
| % ever smokers | **33.1** | **46.0** | **80.5** | **22.6** | **64.1** | **72.5** | **48.3%** |  |
| Family history of CVD (%) | **-** | **58.2** | **-** | **67.6** | **5.3** | **63.0** | **51.6** |  |
| Townsend score |  | **0.355 (3.21)** |  |  |  |  | **-0.21 (0.91)** |  |
| Body mass index (kg/m2) | **25.4 (2.9)** | **27.4 (4.8)** | **26.5 (3.6)** | **25.2 (3.6)** | **27.3 (4.2)** | **27.2 (4.5)** | **25.1 (3.5)** |  |
| Total cholesterol (mmol/L) | **6.36 (1.04)** | **6.79 (1.21)** | **5.61 (0.98)** | **7.08 (1.32)** | **6.12 (1.21)** | **6.12 (1.05)** | **6.44 (1.16)** |  |
| Systolic blood pressure (mmHg) | **144.1 (20.2)** | **154.0 (27.0)** | **146.0 (22.3)** | **144.6 (25.1)** | **144.6 (19.6)** | **137.7 (21.1)** | **121.3 (14.0)** |  |
| Treated hypertension | **2.2%** | **28.6%** | **-** | **16.8%** | **36.0%** | **11.6%** | **5.8%** |  |
| Calendar years for baseline data collection | **1978-80** | **1999-2001** | **1984-88** | **1987-88** | **2004-5** | **1999** | **1992-1993** |  |

**\*adjusted for nested case-control study design, accounting for sampling fraction of controls**

**+In ELSA, follow up was for 5 years so the observed number of events was doubled for the 10 year rate**

**Table 2.** Area under the receiver operating characteristic curve( 95% CI) and detection rates for the combined data

|  |  |  |  |
| --- | --- | --- | --- |
|  | AUROC for combined studies | Detection rate for 5% False positive | Detection rate for 10% False positive |
| Externally weighted gene score | 0.524 (0.508-0.541) | 6.8%  (5.5 to 8.1) | 13.1%  (11.3 to 14.8) |
| QRISK-2 | 0.635 (0.619-0.650) | 11.9%  (10.3 to 13.6) | 21.2%  (19.1 to 23.3) |
| QRISK-2+  Externally weighted gene score | 0.623 (0.608-0.639)  P=0.06\* | 12.0%  (10.3 to 13.6) | 19.6%  (17.5 to 21.6) |

\*P value derived from the comparison with QRISK-2 alone, estimated difference (95% CI) = -0.008 (-0.017 to 0.000).

**Table 3a**: Net reclassification index based on addition of gene score to QRISK, calculated using 10% risk cut off.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of people | |  |  |  |
|  | QRISK + externally weighted gene score  NO CVD (N=15928.64+) | | Reclassified | |  |
| Predicted risk  QRISK | <10 | >=10 | Increased risk | Decreased risk | Net correctly reclassified |
| <10  >=10 | 5475.02  1156.07 | 785.56  8511.99 | 785.56 | 1156.07 | 2.33% (1.8 to 2.9) |
| Predicted risk  QRISK | QRISK + externally weighted gene score  CVD (N=1697.81+) | |  |  |  |
|  | <10 | >=10 |  |  |  |
| <10  >=10 | 352.27  100.58 | 65.36  1179.60 | 65.36 | 100.58 | -2.07 (-3.56 to -0.59) |
| NRI (95% CI)\*  NRI (95% CI)\*\* | 0.25% (-1.33 to 1.83) p=0.76  1.18% (-0.23 to 2.60) p=0.10 | |  |  |  |

\*no adjustment for study

\*\*results from meta-analysis of individual study results (fixed effects)

+ Numbers inflated due to extra weighting assigned to three studies where samples of controls were taken (see Statistical Analysis section)

**Table 3b**: Net-reclassification index based on addition of gene score to QRISK, calculated using 20% risk cut off.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of people | |  |  |  |
|  | QRISK + externally weighted gene score  NO CVD (N=15928.64+) | | Reclassified | |  |
| Predicted risk  QRISK | <20 | >=20 | Increased risk | Decreased risk | Net correctly reclassified |
| <20  >=20 | 9789.24  927.36 | 783.85  4428.19 | 783.85 | 927.36 | 0.90% (0.39 to 1.41) |
| Predicted risk  QRISK | QRISK + externally weighted gene score  CVD (N=1707.7+) | |  |  |  |
|  | <20 | >=20 |  |  |  |
| <20  >=20 | 605.5  124.3 | 146.1  831.9 | 146.1 | 124.3 | -0.25% (-2.09 to 1.58) |
| NRI (95% CI)\*  NRI (95% CI)\*\* | 0.65% (-1.26 to 2.55) p=0.51  0.68% (-1.16 to 2.52) p=0.47 | |  |  |  |

\*no adjustment for study

\*\*results from meta-analysis of individual study results (fixed effects)

+ Numbers inflated due to extra weighting assigned to three studies where samples of controls were taken (see Statistical Analysis section)