ARTICLE OPEN

Check for updates

Elucidating the causal role of age of menarche, adiposity, lipid fractions, and blood pressure upon cardiovascular disease: a multivariable Mendelian randomization study

Yongho Jee[1,6]✉, Wes Spiller[2,3,6], Eleanor Sanderson[2,3], Kate Tilling[2,3], Tom Palmer[2,3], Eunhee Ha[4] and YoungJu Kim[5]

© The Author(s) 2025

This study explores the causal role of multiple correlated risk factors in coronary heart disease (CHD) and ischemic stroke, using Mendelian randomization (MR) analyses with GWAS summary data from both prevalent and incident stroke cases. Thirteen candidate risk factors were considered, including age at menarche, adiposity, lipid fractions, blood pressure, and smoking. Univariable MR identified seven exposures significantly associated with CHD risk, including BMI, blood pressure, LDL, triglycerides, type-II diabetes, and smoking. Notably, HDL showed a protective effect (OR = 0.77, 95% CI: 0.72–0.83), while type-II diabetes was positively associated with CHD (OR = 1.10, 95% CI: 1.05–1.16). For ischemic stroke subtypes, diastolic and systolic blood pressure showed consistent effects across both small vessel and large artery stroke (e.g., DBP OR = 2.27, 95% CI: 1.77–2.89 for small vessel stroke), and HDL again demonstrated protective effects. Multivariable MR (MVMR) further confirmed these associations, though estimates were attenuated. In summary, both univariable and MVMR analyses identified robust associations of lipid fractions and blood pressure with cardiovascular outcomes, highlighting their importance in CHD and ischemic stroke risk across multiple stroke subtypes.

Journal of Human Hypertension (2025) 39:723–734; https://doi.org/10.1038/s41371-025-01048-y

INTRODUCTION

Cardiovascular disease (CVD) is a leading contributor to global mortality, responsible for more than 30% of global deaths [1]. As a consequence, identifying modifiable risk factors for CVD related illness has come to represent a global imperative, facilitating the development of strategies for treatment and prevention [1, 2]. As CVD encompasses a wide range of illnesses with potentially diverse aetiologies, such a goal is best met through an improved understanding of the causal mechanisms underlying each disease.

With the proliferation of genome-wide association studies (GWAS), summary Mendelian randomization (MR) has emerged as a popular methodological framework for estimating causal effects [3–6]. Summary MR methods utilize genetic variant-phenotype association estimates from publicly available GWAS, producing causal effect estimates which are potentially robust to reverse causation or confounding bias [4, 7–9]. While such analyses have often focused on estimating the total effect of a single risk factor on an outcome, the recent development of multivariable summary MR approaches (MVMR) has facilitated evaluation of the conditional effects of multiple exposures simultaneously [10–12]. MVMR approaches are particularly useful when evaluating correlated exposures, providing a more detailed depiction of the causal mechanisms underpinning a given disease.

Previous MR studies have identified several individual risk factors—including BMI, LDL cholesterol, systolic blood pressure, and smoking—as being causally related to CVD outcomes, including coronary heart disease and stroke [13–15]. However, these studies often evaluated each exposure in isolation and did not account for the correlation between multiple risk factors. Recent applications of multivariable MR (MVMR) have begun to address this limitation [10–12], but most studies have been limited to a small number of exposures and have not rigorously addressed the trade-off between model complexity and instrument strength.

In addition, existing studies often rely on summary statistics from prevalent case datasets (e.g., MEGASTROKE), which can introduce selection bias. There is limited evidence comparing results obtained using incident versus prevalent cases in MR analyses, particularly for stroke subtypes such as small vessel and large artery ischemic stroke.

This study addresses these limitations by (i) employing a novel forward selection algorithm to identify the optimal combination of exposures with adequate instrument strength for inclusion in MVMR models, and (ii) systematically comparing causal estimates derived from both prevalent and incident stroke datasets, thereby enhancing our understanding of the methodological and clinical implications of MR-based inference in cerebrovascular disease research.

[1]Advanced Biomedical Research Institute, Ewha Womans University Seoul Hospital, Seoul, Republic of Korea. [2]Population Health Sciences, University of Bristol, Bristol, UK. [3]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [4]Department of Environmental Medicine, Ewha Womans University, College of Medicine, Seoul, Republic of Korea. [5]Department of Obstetrics and Gynecology, Ewha Mokdong Hospital, Ewha Womans University, Seoul, Republic of Korea. [6]These authors contributed equally: Yongho Jee, Wes Spiller. ✉email: jyongho@ewha.ac.kr

Received: 21 October 2024 Revised: 25 June 2025 Accepted: 10 July 2025
Published online: 19 August 2025

This study explores the extent to which the causal effects of risk factors for CHD and ischemic stroke are driven by their relationship with additional correlated exposures. Initially, we identify a set of candidate risk factors through literature review as a means of justifying their inclusion in downstream analyses. Causal effects are then estimated for each exposure, with a view to minimizing the possibility of selectively reporting results only for risk factors satisfying an arbitrary significance threshold. Following the literature review, we adopted age of menarche, blood pressure, adiposity, lipid fractions, physical activity, type-II diabetes, smoking, alcohol consumption, sleep duration, and kidney function as risk factors for which genetic determinants were plausible.

While it was possible to evaluate the effect of each exposure individually, MVMR estimates could not be reliably obtained when fitting all exposures within a single MVMR model due to insufficient conditional instrument strength. This highlights a common issue when performing MVMR analyses, specifically how to select and justify an appropriate MVMR model when all exposures cannot be used simultaneously. In this paper we propose a novel forward selection algorithm maximizing conditional F-statistics for instrument strength to select the greatest feasible number of exposures (see Methods). Assuming that the initial set of risk factors is determined with reference to previous research, such as through literature review, our forward selection approach provides a means through which the greatest number of exposures can be included, while minimizing the extent to which exposure selection is arbitrarily determined through researcher biases.

A further issue explored in this paper is potential bias resulting from the use of GWAS summary data from studies using prevalent ischemic stroke cases. Specifically, the MEGASTROKE consortium leverages data from several case-control studies, and as a consequence resulting estimates could be influenced by selection into the study [16]. This could happen in cases where a risk factor is causally related to ischemic stroke severity, such that participants with specific risk factor configurations are more likely to be recruited into the study. For example, if greater levels of alcohol consumption increased ischemic stroke severity, it could be the case individuals with lower levels of alcohol consumption are enriched in the sample of ischemic stroke cases, because increased alcohol consumption increases the extent to which incidents are fatal and only participants surviving an ischemic stroke are included in the study. This can consequently result in reduced alcohol consumption being associated an increased risk of stroke, when it is actually associated with study participation for participants acting as cases.

The extent to which causal effect estimates are driven by the use of GWAS summary data from prevalent ischemic stroke cases is explored by repeating univariable MR and MVMR analyses with GWAS summary data specific to incident ischemic stroke. GWAS summary data is then used in univariable MR and MVMR analyses, comparing resulting estimates to previous analyses using MEGASTROKE GWAS summary data.

METHODS

Identifying candidate exposures through review

Using the PubMed and Web of Science online databases, the search terms “coronary heart disease”, “myocardial infarction”, “stroke”, “ischemic”, “ischemic”,” hemorrhagic”,” hemorrhagic”, and” atherosclerosis” were used, searching for instances wherein any of the terms were used (a Boolean” OR” operator). Author names, article title, year of publication, journal name, and abstracts for each paper were extracted, covering the time period May 2011-May 2021. A three-stage selection process was then conducted, assessing the title, abstract, and main body of each paper sequentially.

Papers were excluded from the review on the basis of five selection criteria:

1. Inaccessible or written in a non-English language.
2. A review or meta-analysis.
3. An animal or cell model study.
4. Unrelated to cardiovascular disease or conducting analyses where controls were defined as having a pre-existing cardiovascular condition, such as hypertension.
5. Of insufficient quality using Joanna Briggs Institute criteria.

A total of 52,931 papers were identified in the initial search, with 2028 papers remaining after exclusion criteria were applied. With this complete, the subset of 2028 relevant papers were evaluated with respect to paper quality using the Joanna Briggs Institute (JBI) critical appraisal tools (available at https://jbi.global/critical-appraisal-tools). A summary of the number of papers excluded by category is provided in Fig. 1 (left), while a summary of papers identifying statistically significant associations between one or more risk factors and CHD or stroke is presented in Fig. 1 (right).

The set of remaining papers can further be subcategorized by outcome and risk factor of interest. In Fig. 1 (right) a summary of the distribution of papers by outcome of interest (CHD or all subtype stroke) is provided, highlighting sets of frequently examined risk factors. For each paper the presence and direction of significant estimated effects (using a threshold of $p < 0.05$ or corrections for multiple testing where applicable) were extracted. Note that for clarity only exposures for which the number of corresponding papers exceeded five are included in Fig. 1. The source documents used for filtering papers are available at https://github.com/WSpiller/CVDMVMR.

Focusing on modifiable risk factors, analyses exploring the role of diastolic blood pressure (DBP) and systolic blood pressure (SBP) on CHD and ischemic stroke were most frequently identified. Systolic and diastolic blood pressure were frequently found to be positively associated with CHD and stroke risk [17–21]. In addition, several studies have suggested that the effect of risk factors such as insulin resistance may vary with changes in blood pressure [22, 23]. Adiposity was also identified as a risk factor for both stroke and CHD, with increases in measures such as body mass index (BMI) and waist circumference being estimated to increase disease risk [24, 25]. Alcohol consumption, sedentary behavior, and smoking have also been linked to increased risk of stroke and CHD events [26–31].

A number of studies have highlighted a possible role of serum lipids in CHD and stroke risk, with differing directions of effect depending on the lipid fraction investigated. Broadly, low-density lipoprotein (LDL) and triglycerides are associated with increases in CHD and stroke risk, while high-density lipoprotein (HDL) is frequently estimated to have a protective effect [32–34]. The relationship between metabolites and CHD or stroke has also received substantial attention, though it is often the case that metabolites are considered to represent potential biomarkers for conventional risk factors rather than serving as determinants of CHD and stroke risk [35, 36]. In addition, there is a growing body of research focused on the role of sex hormones and development as indicators of future CHD and stroke risk [37, 38].

It is interesting to note that competing downstream health conditions such as type-II diabetes and kidney function were identified. Such studies raise questions regarding whether these conditions can adversely impact risk of subsequent CHD or stroke events, potentially in conjunction with traditional CVD risk factors. Finally, a set of risk factors were identified for which GWAS summary data is scarce, including environmental, psychological and social exposures. As a consequence, such risk factors are not formally considered in this study, though analyses focusing on these elements of cardiovascular health represent promising avenues for future research.

Study and variable descriptions

Genetic variant-exposure associations and corresponding standard errors for the majority of candidate risk factors were obtained from the UK Biobank, a cohort study composed of participants aged 40–69 recruited within the United Kingdom [39]. Data for approximately 500,000 individuals was recorded, including anthropometric measures, self-report questionnaires, and biological measures for downstream genotyping [40]. Age of menarche was selected as a proxy variable for hormonal changes in childhood development, recorded in years for female participants. BMI was used as a measure of adiposity, derived using measured of height and weight obtained at baseline. Baseline automated readings for DBP and SBP were used, measured with mmHg units, while LDL, HDL, and triglycerides were recorded in mg/dL from serum measures. In addition, creatinine concentration in urine was used as an indicator of kidney function, measured in mg/dL at baseline. Measures of BMI, DBP, SBP, lipid fractions and creatinine were standardized using a z-score transformation prior to conducting analyses so as for estimates to be interpreted on a standard deviation scale.

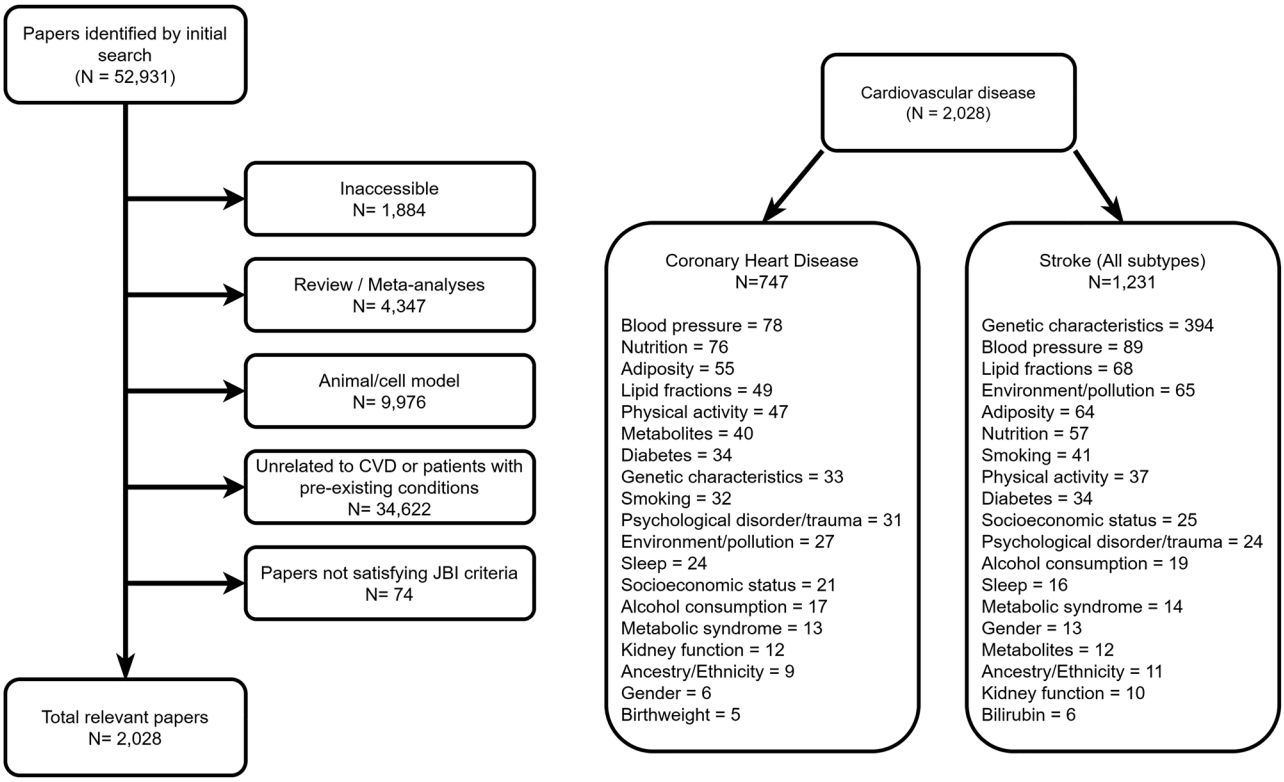


Fig. 1 A panel summarizing the papers initially identified through review (left), and number of exposure specific papers for coronary heart disease (CHD) and stroke (right).

The number of days a participant engaged in 10 min or more of vigorous physical activity was used as a general measure of physical activity, while sleep duration records the average hours of sleep per day for each participant. Average units of alcohol consumed per week, derived by Clarke et al., and excluding former drinkers was used as a measure of alcohol consumption [41], while smoking behaviour is a categorical variable representing past tobacco smoking. Importantly, as categories represent general smoking behaviour tied to frequency of consumption as opposed to average amount consumed, subsequent estimates are interpreted as representing increases in smoking amount, and are not linked with a specific unit increase. Physical activity, sleep duration, alcohol consumption and smoking behaviour measures were obtained using determined using self-report questionnaire at baseline.

Genetic variant-outcome associations for CHD were sourced from the CARDIOGRAMplusC4D consortium [42]. The CARDIOGRAMplusC4D consortium conducted a large-scale GWAS meta-analysis using information from approximately 180,000 participants across a total of 48 studies. CHD was recorded as a binary variable with approximately 60,800 identified cases. Incident CHD cases were defined using International Classification of Diseases 9th Revision (ICD-9) codes 410–414, and ICD-10 codes I20-I25. A sample of mixed ancestry was used, though participants were predominantly of European ancestry (77%). For ischemic stroke summary data were obtained from the MEGASTROKE consortium, including large-artery atherosclerosis (4373 cases), cardioembolic (7193 cases), and small vessel (5386 cases) subtypes as defined by the Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria [43]. MEGASTROKE participants were primarily of European ancestry. The study was reviewed and approved by the Ewha Womans University Seoul Hospital Institutional Review Board (IRB File No: SEUMC 2021-08-026).

For analyses restricted to incident stroke, cases were identified using Hospital Episode Statistics (HES) data within the UK Biobank, generating a binary variable indicating one or more stroke events occurring after baseline measures were recorded. Individuals having previously experienced a stroke event were excluded, and stroke cases were identified using ICD-10 codes I61-6363. Resulting estimates represent effect of each exposure on general ischemic stroke as opposed to specific ischemic stroke subtypes.

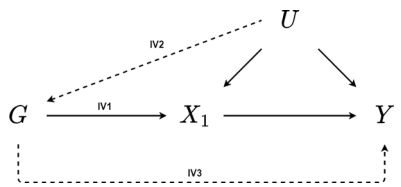


Fig. 2 A directed acyclic graph depicting the assumed relationships for unbiased inference using MR.

Univariable and multivariable Mendelian randomization

MR uses genetic variants, often single nucleotide polymorphisms (SNPs), as instrumental variables to estimate causal effects which are not biased by unmeasured confounding. This is possible in cases where a candidate genetic variant is strongly associated with an exposure of interest (IV1), there are no confounders of the genetic variant and the outcome (IV2), and independent of the outcome when conditioning on the exposure (IV3) [44]. These conditions are illustrated in Fig. 2 for a genetic variant G, exposure X1, outcome Y, and set of unmeasured confounders U.

For an individual genetic instrument, a Wald ratio estimate can be obtained by dividing the genetic variant-outcome association by the genetic variant-exposure association [4, 23]. Where multiple SNPs are available, ratio estimates are often combined through fixed-effects meta-analysis [4, 7]. The inverse variance weighted (IVW) estimate represents an average of multiple ratio estimates, weighted by the inverse of the variance of the genetic variant-outcome association.

Violation of the IV assumptions can introduce bias into causal effect estimates. Assumption IV1 can be tested through an F-statistic, however it is not possible to formally test assumptions IV2 and IV3. However, provided associations violating IV3 are not identically distributed across the range of instruments used, heterogeneity in individual effect estimates can serve as an indicator of IV3 violation [24]. Heterogeneity can be estimated using Cochran's Q-statistic, such that genetic instruments contributing substantially to global heterogeneity are typically identified as invalid [45].

In cases where one or more genetic variants are invalid several methods for obtaining MR estimates that are robust to that violation are available. MR-Egger regression allows for directional pleiotropy by extending the IVW

regression model to include an intercept, where a non-zero intercept is interpreted as an average of pleiotropic effects across the set of genetic variants [7]. This hinges upon an additional assumption that the size of the pleiotropic effect for each SNP is independent of the strength of that SNP as an instrument. The weighted median approach selects the median per SNP ratio estimate with respect to weighting, providing an unbiased causal effect estimate where the weighted proportion of valid genetic variants exceeds 50% [8]. The weighted modal estimator assumes that the modal ratio estimate βj represents estimates obtained from valid genetic variants. This is referred to as the Zero Modal Pleiotropy Assumption (ZEMPA) [9]. This will give an unbiased effect estimate when the largest group of SNPs is not pleiotropic.

MVMR extends the univariable MR framework to contain multiple exposures simultaneously [11, 12], adjusting for each exposure to provide estimates of direct effect. The assumptions of MVMR build upon the existing MR assumptions, with MVMR relying upon genetic variants:

1. Strongly associated with each exposure when conditioning on remaining exposures (MVMR1).
2. No confounding of any of the genetic variants and the outcome (MVMR2)
3. Independent of the outcome when conditioning on all included exposures (MVMR3)

Assumption MVMR1 is assessed by calculating the conditional F-statistic for each exposure, typically using a threshold value of 10 to indicate sufficient instrument strength for conducting analyses [12, 46]. Violations of MVMR3, resulting from continued invalidity after controlling for additional exposures can be evaluated by calculating an adjusted Q-statistic, as described in Sanderson et al. [12]. Crucially, the inclusion of additional exposures has the potential to address the limitations of the univariable summary MR approaches previously discussed. Specifically, comparing univariable and conditional F-statistics can highlight the extent to which genetic variants are predictive for multiple exposures, resulting in potential invalidity.

MVMR model selection

To select an appropriate MVMR model with respect to instrument strength, a novel forward selection algorithm is presented with the aim of including the highest number of exposures for which conditional F-statistics exceed a threshold value of 10. Initially, all possible pairs of exposures are evaluated with respect to their conditional F-statistics. The mean F-statistic for each pair is then calculated, weighted by the inverse of the absolute value of the difference between F-statistics within a given pair. The rationale behind weighting the mean F-statistic for each pair stems from instances where two exposures have a high mean F-statistic, driven primarily by one of the exposures. For example, consider two pairs of exposures A and B where each pair has a mean conditional F-statistic of 50. Further, assume the conditional F-statistics for pair A and pair *B* are ($F1 = 10$, $F_2 = 90$) and ($F1 = 50$, $F_2 = 50$) respectively. In this case, the similarity in mean conditional F-statistic could mistakenly imply that the conditional F-statistics for both pairs are similar.

In the above example the decision to prefer pair B over pair A is arbitrary, given that the conditional F-statistics for all exposures exceeds the commonly used threshold of 10. The distinction is important, however, when considering the inclusion of additional exposures within the MVMR model. A conditional F-statistic reflects the strength of association between a set of genetic instruments and a given exposure, conditioning on additional exposures included within the MVMR model. Consequently, if an exposure is added which serves as a mediator between an existing exposure and one or more of its corresponding genetic instruments, it would be follow that the conditional F-statistic for the existing exposure would decrease. Essentially, a proportion of the instrument strength for the existing exposure can be attributed to the relationship between the two exposures, and is removed when conditioning on the additional exposure. Returning to hypothetical exposure pairs A and B, pair B can more easily accommodate reductions in F-statistic while satisfying the threshold of 10 in contrast with pair A.

Before continuing it is important to highlight two key considerations when using conditional F-statistics for MVMR model selection. First, the inclusion of an additional exposure does not necessarily imply conditional F-statistics for the existing exposures will decrease. Conditional F-statistics are most likely to decrease when the additional exposure lies on the causal pathway from the genetic variants to an existing exposure. Second, when conditional F-statistics are found to decrease, and the candidate additional exposure is omitted from the model, resulting direct effect estimates for the remaining exposures implicitly include the effect of the omitted exposure on the outcome. If the effect of the omitted exposure is entirely mediated through the existing exposures, or the effect of the existing exposures is mediated entirely through the omitted exposure, then omitting the additional exposure will not introduce bias into MR estimates. However, where the omitted exposure exerts a direct effect on the outcome the resulting MR estimates will be biased in the direction of this association. These two scenarios describe vertical and horizontal pleiotropic relationships respectively.

Statistical analysis

For MR analyses sets of independent genetic variants were initially for each exposure using the MR-Base online GWAS database [47, 48]. A p-value threshold of $p < 5 \times 10-8$ was used to identify genetic variants robustly associated with the phenotype of interest, and linkage-disequilibrium (LD) pruning was implemented using a threshold of $r2 < 0.01$ and a clumping window of 10,000 kb to ensure only independent genetic variants were selected. For each outcome univariable MR analyses were performed, specifically using IVW, MR-Egger, weighted median, and weighted modal approaches.

To select an appropriate MVMR model we initially calculated the weighted mean conditional F-statistic for each exposure pair. With this complete, we selected the pair with the highest weighted mean conditional F-statistic and constructed a set of models by including one of the remaining exposures individually. This process was iterated until remaining exposures either failed to exceed the given threshold, or the inclusion of an additional exposure would lower the conditional F-statistic for any included exposure below the instrument strength threshold. Ideally it would be desirable to include all potentially relevant exposures within a single MVMR model, yet it was not possible to do so while maintaining sufficient conditional instrument strength. The final model used in MVMR analyses therefore reflects a trade-off between including modifiable risk factors of interest and producing causal estimates of sufficient precision so as to be informative. GWAS association estimates were obtained from the MR-base GWAS database, and analyses were performed using the TwoSampleMR, RadialMR, and MVMR R packages [45, 48]. Analyses were conducted in R (version 4.0.3), and code for the analyses is available at https://github.com/WSpiller/CVDMVMR.

To evaluate potential bias from using data derived from studies using prevalent stroke cases, we conducted a GWAS using genetic and HES data from UK Biobank using information on unrelated participants of European ancestry. Outcome data was sourced from the HES database, extracting date of hospital admission and corresponding diagnoses categorised using ICD-10 codes. Initially participants with a history of ischemic stroke prior to recruitment were removed, corresponding to ICD-10 codes I61–I63. Individuals experiencing a stroke event after followup were then categorised as cases ($N = 9527$). Finally, a random subset of individuals was selected as controls, maintaining a case to control ratio of (1:10) and a total sample size $N = 95,270$. Effects were obtained using BOLT-LMM adjusting for participant age and sex in addition to array batch effects. Resulting estimates and standard errors were transformed so as to be interpreted as a log-odds ratio, and descriptive statistics for the GWAS sample are provided in Supplementary Table S1.

RESULTS

In univariable MR analyses thirteen sets of genetic instruments were obtained, corresponding to each of the exposures previously described. A total of 2038 SNPs were used across all exposures, with mean F-statistics across all analyses ranging from 36–177. Summary data for each univariable analysis available at https://github.com/WSpiller/CVDMVMR, in addition to code for downloading and replicating each analysis. Fig. 3 shows a forest plot displaying the IVW estimate and 95% confidence interval for each exposure, grouped by outcome. Effect estimates have been exponentiated so as to be interpreted on an odds ratio scale. To account for multiple comparisons, a Bonferroni correction was applied, resulting in a significance threshold of $p < 0.0038$ (0.05/13).

For CHD seven exposures showed evidence of a positive effect, including BMI, DBP, SBP, LDL, triglycerides, type-II diabetes and smoking. BMI, blood pressure and LDL exert effects of similar magnitude, representing change in CHD risk per standard

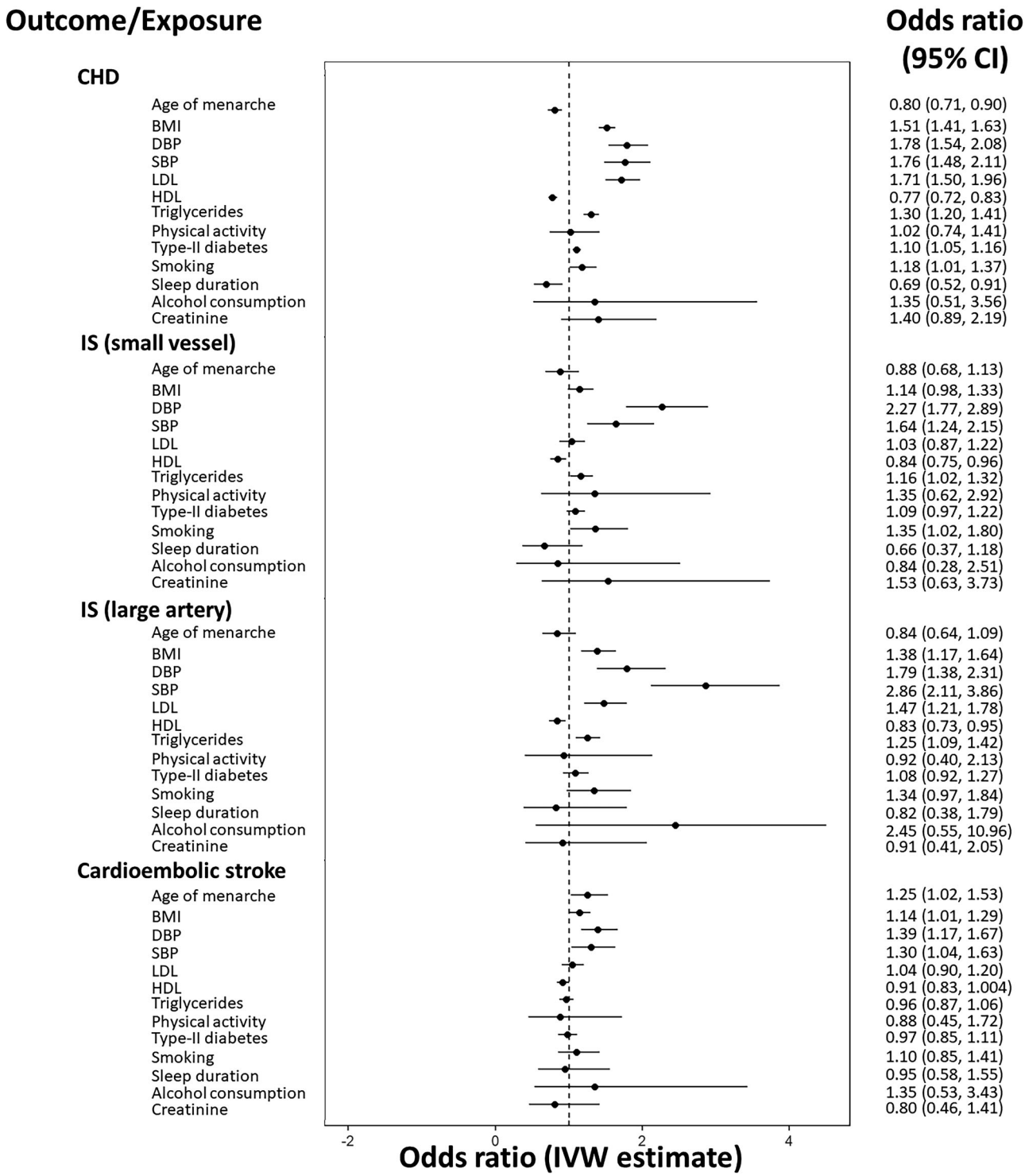


Fig. 3 A forest plot showing IVW univariable MR estimates. Estimates are grouped by outcome. Effect estimates on the odds ratio scale, with accompanying 95% confidence intervals are illustrated, with corresponding values shown to the right of the plot.

deviation increase. Genetically predicted type-II diabetes status was found to increase the risk of CHD by 10% (OR = 1.10, 95% CI = 1.05, 1.16), while increased frequency of tobacco consumption was positively associated with CHD risk (OR = 1.18, 95% CI = 1.01, 1.37). Later onset of menarche (OR = 0.80, 95% CI = 0.71, 0.90), HDL (OR = 0.77, 95% CI = 0.72, 0.83) and increased sleep duration (OR = 0.69, 95% CI = 0.52, 0.91) were found to be protective with respect to CHD. Estimates for increased concentration of creatinine, representing reduced kidney function, were suggestive of a positive effect on CHD risk for which the analyses was under-powered (OR = 1.40, 95% CI = 0.89, 2.19). Similarly, the wide confidence interval for alcohol consumption reflects a lack of statistical power to detect a causal effect on CHD. Physical activity showed no evidence of a causal effect on CHD (OR = 1.02, 95% CI = 0.74, 1.41).

Considering small vessel ischemic stroke DBP (OR = 2.27, 95% CI = 1.77, 2.89) and SBP (OR = 1.64, 95% CI = 1.24, 2.15) appear to have the most substantial effects. Triglycerides, smoking

behaviour showed evidence of positive associations with small vessel ischemic stroke risk, while HDL appeared to have a protective effect (OR = 0.84, 95% CI = 0.75, 0.96). LDL showed no evidence of a directional effect on small vessel ischemic stroke (OR = 1.03, 95% CI = 0.87, 1.22). Sleep duration showed limited evidence of a protective effect for which the analyses were underpowered (OR = 0.66, 95% CI = 0.37, 1.18), while power constraints limited the extent to which physical activity, alcohol consumption or creatinine could be considered.

Comparing small vessel and large artery ischemic stroke subtypes, blood pressure again exerted the greatest effects on large artery ischemic stroke risk. The positive effects of BMI (OR = 1.38, 95% CI = 1.17, 1.64) and triglycerides (OR = 1.25, 95% CI = 1.09, 1.42) on large artery ischemic stroke were greater in magnitude, while smoking behaviour and HDL yielded effects in agreement with small vessel ischemic stroke. Notably, LDL is estimated to have a substantial positive effect on large vessel ischemic stroke (OR = 1.47, 95% CI = 1.21, 1.78), while the effect of type-II diabetes was attenuated (OR = 1.08, 95% CI = 0.92, 1.27). Remaining exposures had insufficient precision to identify causal effects.

For cardioembolic stroke DBP was the only exposure with an observed association after correcting for multiple testing (OR = 1.39, 95% CI = 1.17, 1.67). Estimates for age of menarche show evidence of a positive association with cardioembolic stroke.

(OR = 1.25, 95% CI = 1.02, 1.53) in contrast with the protective effect observed for CHD. Estimates for BMI and SBP seem to indicate potential roles as risk factors, in addition to a protective effect of HDL (OR = 0.91, 95% CI = 0.83, 1.004). LDL, physical activity, sleep duration, tobacco consumption and creatinine showed no evidence of a directional causal effect, while analyses of alcohol consumption lacked statistical power.

Overall, the results depicted in Fig. 3 highlight key differences in potentially relevant risk factors across the range of outcomes considered. Age of menarche seems related to CHD and cardioembolic stroke, though the direction of causal effect differs. BMI and blood pressure are consistently highlighted as risk factors exerting a positive effect on all outcomes, while LDL and triglycerides appear to only influence CHD and large artery ischemic stroke. HDL is estimated to have a protective effect with respect to the range of outcomes, while a positive effect of type-II diabetes was only observed for CHD. Smoking was a risk factor for all outcomes except for cardioembolic stroke, while estimates indicative of a causal role of sleep duration and kidney function were only found for CHD. Unfortunately, analyses of alcohol consumption and physical activity were not sufficiently precise to identify potential directional effects.

The sets of univariable estimates presented were obtained using the IVW method, and as a consequence, can be biased where horizontal pleiotropic pathways are present (violations of IV3). Using Cochran's Q-statistic as an indicator of potential pleiotropic bias, substantial heterogeneity was only present for age of menarche when CHD served as an outcome (Q = 205.7, $p < 0.001$). Heterogeneity appeared to be present in all BMI and blood pressure analyses, and for lipid fractions all analyses showed evidence of heterogeneity with the exception of HDL (Q = 313.49, $p = 0.561$) and triglycerides (Q = 269.43, $p = 0.600$) in relation to cardioembolic stroke. Analyses focusing on physical activity and type-II diabetes did not show evidence of substantial heterogeneity, while smoking (Q = 156.8, $p < 0.001$) and alcohol consumption (Q = 14.9, $p < 0.001$) only exhibited heterogeneity when considering CHD. Creatinine had evidence of heterogeneity in analyses of CHD (Q = 47.95, $p < 0.001$) and small vessel ischemic stroke (Q = 35.59, $p = 0.027$). Heterogeneity statistics for each analysis are provided in Supplementary Table S2.

In cases where heterogeneity was observed, invalid instruments were identified and removed based on their contribution to global heterogeneity. This was achieved by evaluating the Q-statistic for each instrument individually, using a p-value threshold with a multiple testing correction equal to 0.05 divided by the total number of instruments in the IVW analysis. IVW analyses performed using sets of instruments following outlier removal, in addition to applying sensitivity analyses, produced results in agreement with the IVW estimates presented in Fig. 3 (see Supplementary Material Table S3).

In MVMR analyses it was not possible to fit all 13 exposures simultaneously while maintaining sufficient conditional instrument strength to estimate causal effects. For example, fitting a six exposure MVMR model yielded conditional F-statistics for age of menarche ($F \approx 11$), BMI ($F \approx 6$), SBP ($F \approx 9$), LDL ($F \approx 4$), HDL ($F \approx 2$), or triglycerides ($F \approx 2$). As previously discussed, the attenuation of conditional F-statistics when fitting additional exposures could be due to either a low proportion of the total set of SNPs being associated with a particular exposure or because the set of genetic variants is predictive of multiple exposures (see Methods) [12].

As a consequence, effect estimation using MVMR required a subset of exposures to be used. Using the forward selection procedure previously described (see Methods), all potential pairs of exposures were evaluated, with HDL and triglycerides providing the highest weighted sum of conditional F-statistics. After evaluating additional exposures sequentially LDL and SBP were also included in a four exposure MVMR model. Fig. 4 is a heat map presenting the conditional F-statistic for each exposure (labelled exposure 1), and illustrating the magnitude of the weighted mean conditional F-statistic as a gradient. The conditional F-statistic corresponds to the exposure labelled on the x-axis when fit in a two variable MVMR model including the corresponding exposure on the y-axis. A colour gradient is used to indicate the magnitude of the weighted mean conditional F-statistic. When including LDL, HDL, triglycerides, and SBP in the MVMR model the number of instruments differed across the range of outcomes. This reflects differences in the availability of outcome-specific GWAS summary data for specific SNPs across selected GWAS. Specifically, a total of 1261 instruments were used to estimate effects on CHD, 1296 for small vessel ischemic stroke, 1300 for large artery ischemic stroke and 1291 for cardioembolic stroke. Fig. 5 shows the estimates obtained using all exposures simultaneously, with superimposed univariable IVW estimates for ease of comparison.

When adjusting for multiple exposures, LDL (OR = 1.71, 95% CI = 1.50, 1.96), triglycerides (OR = 1.12, 95% CI = 1.03, 1.22), and SBP (OR = 1.75, 95% CI = 1.66, 1.84) were positively associated with CHD. The effects of LDL and triglycerides were attenuated but directionally consistent with univariable MR estimates. HDL and SBP were found to be in agreement with univariable MR estimates, with a substantial improvement in precision for SBP. The observed agreement in effect estimates for SBP suggests previous univariable estimates are unlikely to be driven by genetic effects shared between exposures included in the MVMR model, though the attenuation visible for LDL and triglycerides suggests that there may be some overlap in genetic effects in univariable analyses. This could be interpreted as an instance where effects are greater due to horizontal pleiotropic effects.

For small vessel ischemic stroke, a protective effect of HDL similar in magnitude to the corresponding univariable MR estimate was observed (OR = 0.81, 95% CI = 0.64, 0.98). SBP was again estimated to have a positive effect on small vessel ischemic stroke (OR = 1.75, 95% CI = 1.58, 1.93), while the effect of LDL remained negligible (OR = 0.94, 95% CI = 0.80, 1.08). Comparing the univariable and MVMR estimated effects of triglycerides, the positive association with small vessel ischemic stroke is not observed when adjusting for LDL, HDL and SBP (OR = 1.01, 95% CI = 0.84, 1.19). This would suggest that the positive association observed in the univariable analyses may be driven by associations with the additional exposures included in the MVMR model.

Considering large artery ischemic stroke estimates for HDL (OR = 0.79, 95% CI = 0.60, 0.97) and SBP (OR = 2.34, 95% CI = 2.15, 2.52) remain directionally consistent with the corresponding

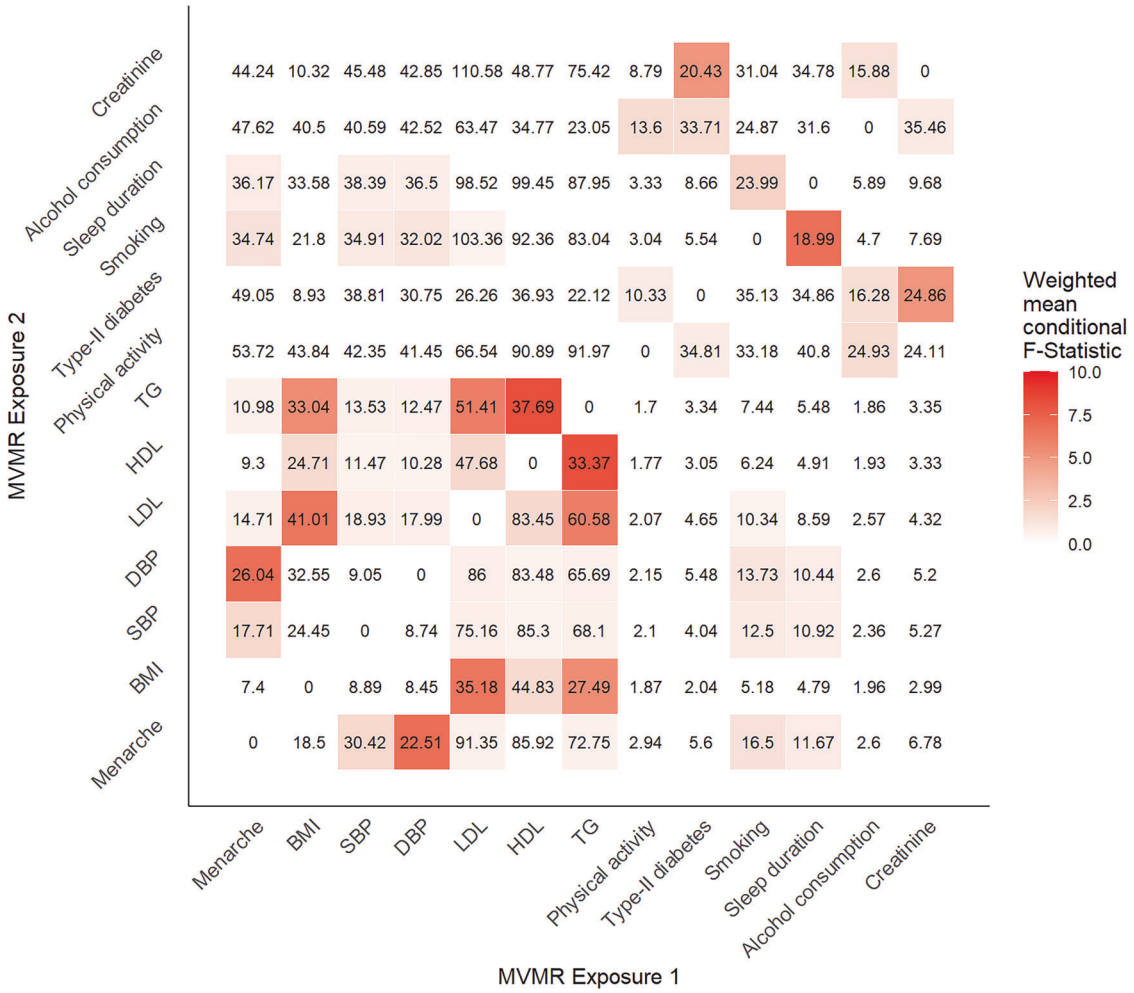


Fig. 4 A heatmap showing the conditional F-statistic for each exposure when included in a two variable MVMR model. The conditional F-statistic corresponds to the exposure labelled on the x-axis when fit in a two variable MVMR model including the corresponding exposure on the y-axis. A colour gradient is used to indicate the magnitude of the weighted mean conditional F-statistic.

univariable MR estimates. SBP continues to have a more substantial effect on large artery ischemic stroke in comparison with small vessel ischemic stroke, and a substantial increase in precision is observed for both stroke subtypes. When adjusting for additional exposures in the MVMR model, the positive associations of triglycerides and LDL are greatly attenuated to include the null with respect to large artery ischemic stroke (OR = 1.11, 95% CI = 0.95, 1.27).

Finally, estimates in relation to cardioembolic stroke were largely in agreement with previous univariable MR estimates. Evidence of a protective effect of HDL is observed (OR = 0.87, 95% CI = 0.72, 1.01), while the positive effect of SBP on cardioembolic stroke risk remains similar in magnitude to the univariable MR estimate (OR = 1.30, 95% CI = 1.15, 1.44).

Analyses using incident ischemic stroke from UK Biobank

After conducting a GWAS of incident stroke using individual-level data from the UK Biobank, univariable MR and MVMR analyses were performed using GWAS summary data specific to incident stroke. Fig. 6 shows resulting estimates, including MVMR estimates using LDL, HDL, triglycerides and SBP. Comparisons with prevalent ischemic stroke are made specifically with reference to small vessel and large artery ischemic stroke subtypes. A subset of GWAS summary data for incident stroke, including all variants selected as instruments in univariable MR and MVMR analyses is available online at https://github.com/WSpiller/CVDMVMR.

Evaluating univariable MR and MVMR estimates in relation to incident ischemic stroke LDL, HDL, triglycerides, and SBP remain directionally consistent. As was the case for prevalent stroke, the precision of the estimated effect of SBP is substantially improved (OR = 1.52, 95% CI = 1.41, 1.63), while LDL and triglycerides exhibit negligible effects. Further, the protective effect of HDL remains present after adjustment for additional exposures (OR = 0.86, 95% CI = 0.76, 0.97).

When comparing estimates using prevalent and incident stroke summary data an initial observation is the more pronounced protective effect of age of menarche on ischemic stroke risk (OR = 0.86, 95% CI = 0.74, 1.001). BMI, DBP, and SBP are again identified as risk factors for incident ischemic stroke, with blood pressure again exerting the most substantial effect. The protective effect of HDL (OR = 0.86, 95% CI = 0.76,0.97) on incident ischemic stroke remains consistent analyses using prevalent stroke data, as well as the positive effect of smoking behaviour (OR = 1.15, 95% CI = 0.94, 1.40).

A notable difference between the incident stroke and prevalent stroke analyses is the estimated effects of LDL and triglycerides. Using prevalent stroke data, a positive effect of triglycerides on small vessel and large artery stroke was observed, while a positive effect of LDL was only observed for large artery ischemic stroke. Assuming the univariable MR estimate for LDL is specific to large artery stroke, one explanation could be that the general ischemic stroke measure used for incident cases is enriched for small vessel

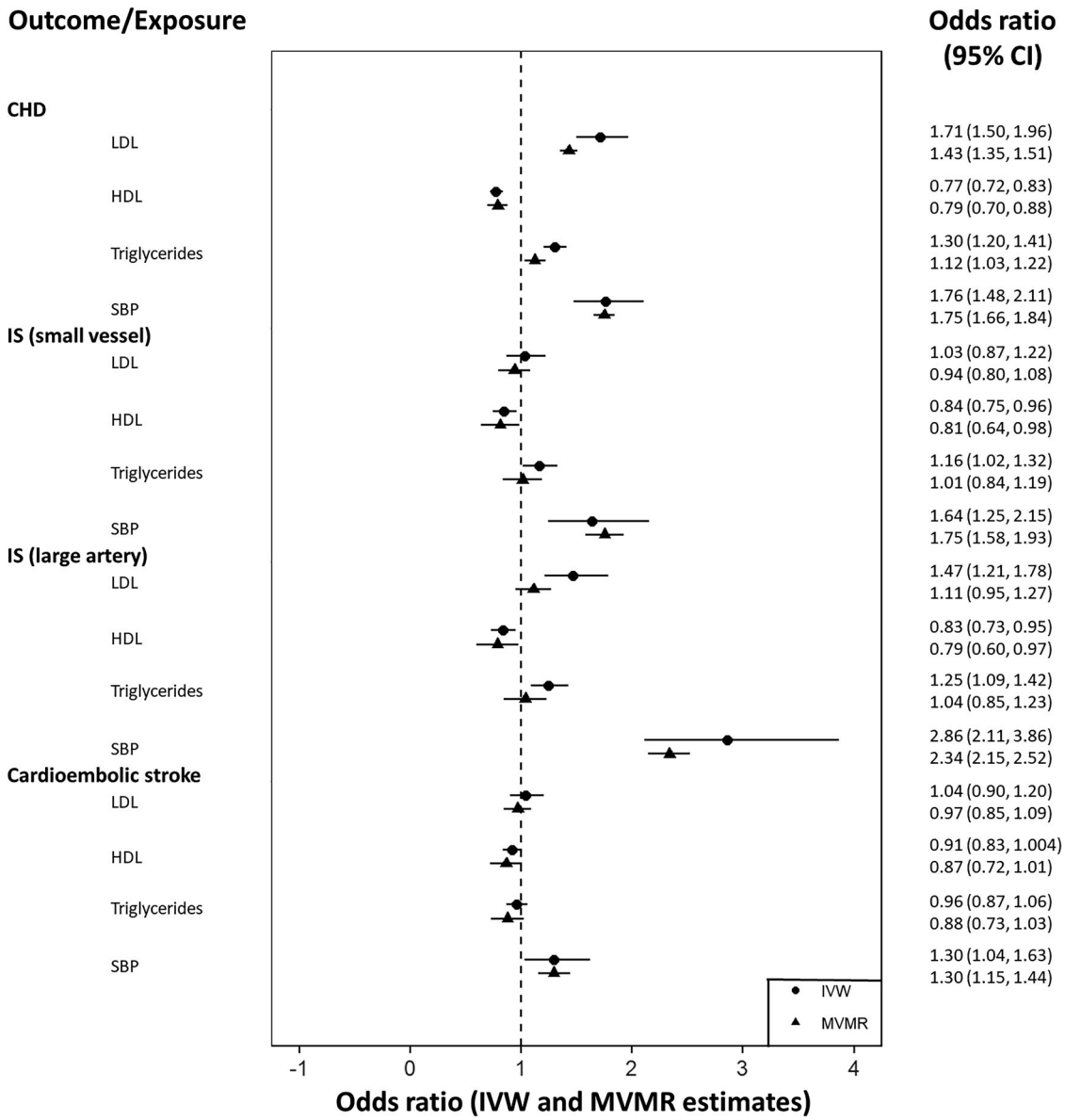


Fig. 5 A forest plot showing MVMR MR estimates. Estimates are grouped by outcome, representing the direct effect of each exposure including all exposures within each MVMR model. Effect estimates on the odds ratio scale, with accompanying 95% confidence intervals are illustrated, with corresponding values shown to the right of the plot. Univariable MR estimates for each exposure are included for comparison.

ischemic stroke. This would have the effect of producing an average causal effect weighted by the ratio of ischemic stroke subtypes included in the general measure of ischemic stroke. This could also explain the discrepancy with triglycerides, where the effect is specific to stroke subtypes within the general measure of incident ischemic stroke. However, when considering MVMR estimates the effects of LDL and triglycerides appear negligible. If the observed associations from univariable MR analyses are driven by associations with exposures included in the MVMR model, then estimates subject to ischemic stroke subtype-specific pleiotropic bias are likely of limited utility in practice. The broad agreement in findings using incident and prevalent ischemic stroke data suggests that this feature of the MR analyses is not driving observed causal effects.

DISCUSSION

This study has explored the potential role of thirteen candidate risk factors on CHD and subtypes of ischemic stroke. Evaluating the role of age of menarche using univariable summary MR, there was evidence of a negative association with CHD, such that menarche occurring at an early age was protective against CHD. This is interesting given that age of menarche is positively associated with relative body size and rate of aging, potential risk factors for CHD [49]. As highlighted by review, such an association could result from fluctuations in hormone levels, in a similar fashion to links between CHD and menopause. One potential explanation for this could be reduced exposure to biological changes related to childhood development, for example exposure to ovarian oestrogen, over the life course. However, while age of menarche is commonly used as a proxy for hormonal timing, it may not fully reflect the complex and heterogeneous endocrine changes that occur during early development. These hormonal trajectories are influenced by multiple genetic, nutritional, and environmental factors, which are not captured by a single timing marker. When using an MVMR model, instrument strength was insufficient to evaluate causal effects, particularly for age of menarche where the sample size for the GWAS was smaller due to

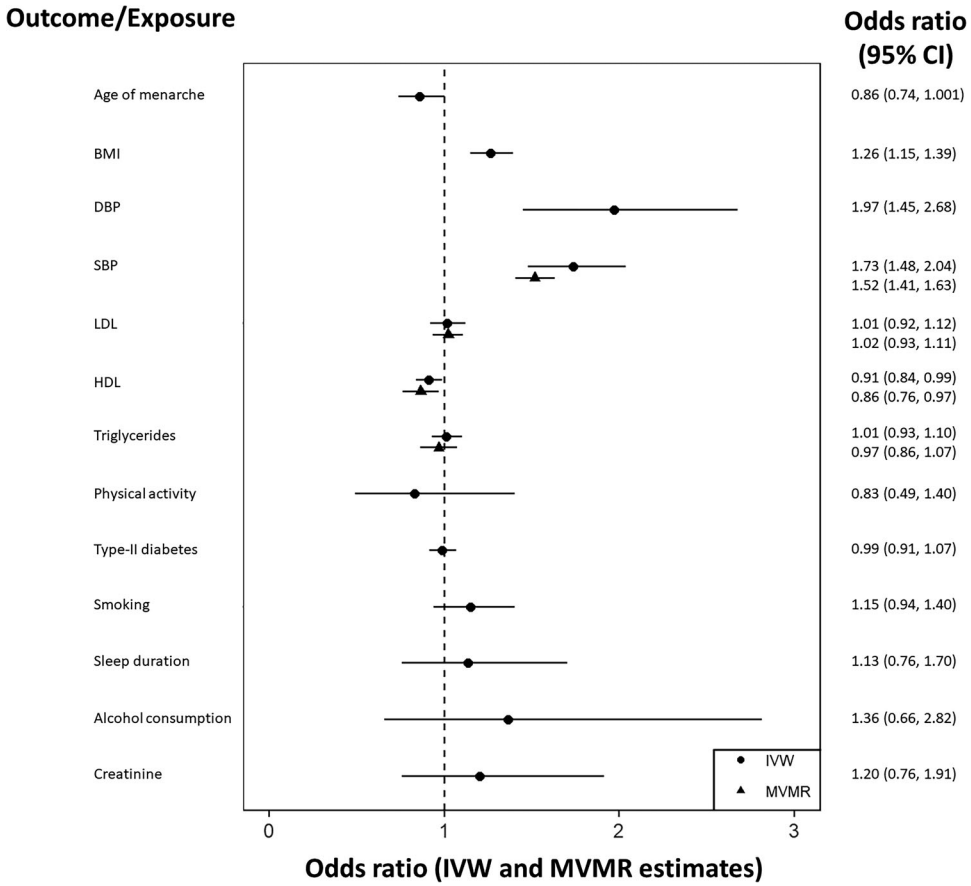


Fig. 6 A forest plot showing IVW univariable and MVMR MR estimates for incident stroke using data from the UK Biobank. Effect estimates on the odds ratio scale, with accompanying 95% confidence intervals are illustrated, with corresponding values shown to the right of the plot.

analyses only including female participants. This implies that included risk factors are linked to age of menarche through shared genetic pathways.Similarly, BMI was identified in univariable analyses as a risk factor for CHD and large artery ischemic stroke, yet was not included in the MVMR model due to insufficient conditional instrument strength. As such, if age of menarche or BMI exert direct effects upon CHD independent of lipid fractions or SBP MVMR estimates could exhibit pleiotropic bias through horizontal pleiotropy. It is, however, also possible that age of menarche and BMI lay on the causal pathway between lipid fractions or SBP and the outcomes considered, though this does not seem biologically plausible as it seems unlikely SBP would be causally upstream of BMI. Such a scenario, termed vertical pleiotropy, would not necessarily result in biased effect estimates.

Blood pressure was consistently found to have a positive causal effect on all outcomes across univariable and MVMR analyses, echoing findings from previous analyses [50]. MVMR analyses exhibited a notable increase in estimate precision, which can be explained a reduction in univariable MR heterogeneity when adjusting for lipid fractions. Positive effect estimates of blood pressure on all forms of ischemic stroke highlight an area for effective intervention, demonstrated through the effectiveness of blood pressure lowering medication such as ACE inhibitors with respect to stroke prevention [51].

A protective effect of HDL was frequently observed with respect to ischemic stroke, suggesting it may serve as a potential target for ischemic stroke prevention. It is, however, important to note that such effects have not been observed in randomized control trials evaluating the effectiveness of HDL-raising drugs. For example, a systematic review conducted by Kaur et al. highlights an apparent a lack of evidence of a protective effect of HDL with respect to all-cause mortality and secondary outcomes including ischemic stroke [52]. This study also considered the use of prevalent stroke cases in MEGASTROKE as a potential explanation for the observed protective effect of HDL, motivating the use of incident stroke data from the UK Biobank. The protective effect remained consistent when restricting analyses to data using incident stroke cases, suggesting that if HDL is not causally relevant other mechanisms may be responsible such as relationships between HDL levels and statin use [53].

Sleep duration was found to have a protective effect on CHD, while the impact of sleep on ischemic stroke subtypes was negligible. It should be noted, however, that the effect of sleep on small vessel ischemic stroke was indicative of a potential protective effect for which statistical power was insufficient. Similarly, increased tobacco consumption had marginal effects on CHD, small vessel ischemic stroke and large artery ischemic stroke. Rather than relying on p-value thresholds alone to highlight causal effects it can in many cases be prudent to consider such estimates with respect to their relative precision. Where indicated, in cases where confidence intervals corresponding to an estimate are noticeably wide, a more conservative interpretation is that there may be an effect which could be identified with improvements in sample size. One example where such an effect seems plausible is the univariable MR estimate for creatinine on CHD, which appears suggestive of a positive effect. It is with these limitations in mind that effects which border classic statistical significance thresholds have been highlighted, rather than relying on p-value thresholds alone [54].

Using summary MR approaches affords this study a number of strengths. Initially, the large study samples of UK Biobank,

CARDIOGRAMplusC4D, and MEGASTROKE afford substantial statistical strength to identify causal effects. It is important to note, however, that large-scale studies with GWAS summary data related to the exposures and outcomes of interest are still relatively limited. As a consequence, it was not possible to identify genetic instruments in an external independent sample, and it is possible that instrument-exposure associations suffer from Winner's curse. This describes a situation wherein observed associations occur due to specific features of a sample, and are subsequently unlikely to be replicated in alternative GWAS. With increases in study sample sizes, future work could leverage improvements in instrument strength to include all risk factors, allowing for the adjustment of potential pleiotropic pathways.

An additional strength of this study is the use of a two-sample study design when conducting summary MR analyses. Such a study design is necessary for the application of sensitivity analyses such as MR-Egger regression, and in the univariable MR setting ensures results more conservative when weak instruments are utilised. Though findings were consistent when applying sensitivity analyses in the univariable MR setting, substantial heterogeneity served as an indicator for potential violations of the MR assumptions, warranting further examination using MVMR.

A novel aspect of this work is the use of a novel model selection algorithm for MVMR, implemented by maximising weighted conditional F-statistics across candidate exposures. While an initial set of candidate exposures is best identified with reference to previous research, use of such an approach can prove helpful when selecting a subset of exposures from a range of equally viable risk factors. Code for implementing the method and replicating study findings is available at https://github.com/WSpiller/CVDMVMR. One limitation of this study relates to the source of GWAS summary data for univariable analyses evaluating age of menarche as an exposure. Due to a lack of available data, outcome GWAS were not sex specific, and as a consequence it is possible that the inclusion of male participants attenuates variant-outcome associations towards the null. Essentially, the contribution of a genetic instrument to the outcome of interest through age of menarche will necessarily not be present in male participants. Use of GWAS summary data including male and female participants has relied on adjustment for participant sex in the GWAS to account for differences in the effect of genetic variants on their respective traits. For example, in the GWAS of BMI, adjustment for participant sex essentially removes the contribution of participant sex to the association estimate. With increases in sample size and available GWAS summary data an evaluation of the impact of using sex-specific data represents a promising area for future research. Several limitations should be considered when interpreting our findings. First, continuous variables such as BMI and blood pressure were standardized using z-scores to facilitate comparability across exposures. While this is a common approach in MR studies, it may reduce clinical interpretability by expressing effects per standard deviation rather than absolute units. Similarly, ordinal variables such as smoking behavior were treated as continuous to maintain analytical consistency across exposures. This simplification is widely adopted in MR literature, but may introduce bias if the exposure–outcome relationship is non-linear or threshold-dependent.

Second, the selection of exposures in the MVMR model was constrained by the strength of the genetic instruments, as determined by conditional F-statistics. While this data-driven approach helps to avoid weak instrument bias, it may inadvertently exclude biologically meaningful exposures that fail to meet statistical thresholds. Such exclusion risks overlooking important components of disease etiology, particularly in cases where weak instruments reflect limited GWAS power rather than a lack of causal relevance. Future work incorporating larger datasets or alternative model selection strategies may help recover these omitted pathways.

Third, as with all MR studies, the validity of causal estimates depends on three key instrumental variable assumptions: that the instruments are strongly associated with the exposure (IV1), not associated with confounders (IV2), and affect the outcome only through the exposure (IV3). While IV1 can be empirically tested using F-statistics, IV2 and IV3 are generally untestable. To mitigate this, we employed MR-Egger, weighted median, and leave-one-out sensitivity analyses, which provide some indication of robustness to assumption violations. However, these methods rely on additional assumptions that are also not empirically verifiable. For instance, MR-Egger depends on the InSIDE (Instrument Strength Independent of Direct Effect) assumption, which cannot be formally tested. Thus, while these sensitivity methods enhance confidence in the findings, residual uncertainty due to untestable assumptions cannot be fully eliminated.A final limitation of this study is the use of a general ischemic stroke measure when conducting analyses of incident stroke. This was motivated by a lack of ischemic stroke subtype-specific cases with which to conduct GWAS and implement downstream summary MR approaches. With the continued proliferation of large-scale population-based cohorts with genetic data and hospital record linkage, it is hoped that these analyses can be revisited with a view to identifying unique aspects of the causal mechanisms underpinning different subtypes of ischemic stroke.

Our last important limitation of our study is the generalizability of the findings to populations beyond those of European ancestry. The data sources for the two-sample Mendelian Randomization (MR) analyses we conducted were primarily based on populations of European descent. This may impact the applicability of our findings to other populations with different genetic backgrounds and environmental exposures. It is well established that genetic associations can vary significantly across different populations due to genetic diversity and varying environmental interactions [55]. For instance, allele frequencies and linkage disequilibrium patterns differ between populations, which can influence the strength and direction of genetic associations observed [56]. Consequently, the causal relationships inferred from MR studies based on European populations may not be directly transferable to non-European populations. To enhance the generalizability of MR findings, future studies should aim to include more diverse populations. This can be achieved by utilizing GWAS data from non-European cohorts and conducting trans-ethnic meta-analyses. Such approaches can help to identify genetic variants that are robust across different populations and to understand population-specific genetic effects [57].

CONCLUSIONS

This study demonstrates that multiple correlated risk factors, including lipid fractions and blood pressure, have significant and directionally consistent effects on coronary heart disease and ischemic stroke, with the univariable and multivariable Mendelian Randomization analyses.

SUMMARY

What is known about the topic?

- Cardiovascular disease (CVD) is a leading contributor to global mortality, responsible for more than 30% of global deaths.
- While numerous epidemiological studies have reported the major risk factors for CVD, the causal effects of these risk factors remain controversial

What Is New?

- This study uses univariable and multivariable Mendelian randomization (MR) to explore the influence of multiple correlated risk factors on both prevalent and incident coronary heart disease (CHD) and ischemic stroke, and introduces a novel forward selection algorithm for multivariable MR (MVMR) analysis.

What Question Should Be Addressed Next?

- Future research should investigate the specific mechanisms through which high-density lipoprotein (HDL) exerts a protective effect on CHD and ischemic stroke, and further validate the findings using diverse populations and datasets.

DATA AVAILABILITY

The genetic association data for coronary heart disease (CHD) were obtained from the CARDIoGRAMplusC4D consortium (http://www.cardiogramplusc4d.org/), and data for ischemic stroke were obtained from the MEGASTROKE consortium (https://www.megastroke.org/). Both datasets are publicly available summary-level genome-wide association study (GWAS) data and contain no individual-level information. No additional data were generated or analyzed in this study.

REFERENCES

1. World Health Organization. Monitoring health for the sustainable development goals. Geneva: World Health Organization; 2020.
2. Stewart J, Manmathan G, Wilkinson P. Primary prevention of cardiovascular disease: a review of contemporary guidance and literature. JRSM Cardiovasc Dis. 2017;1:2048004016687211.
3. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32:1–22.
4. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37:658–65.
5. Sanderson E, Glymour M, Holmes M, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. Nat Rev Methods Primers. 2022;2:6.
6. Richmond RC, Smith GD. Mendelian randomization: concepts and scope. Cold Spring Harb Perspect Med. 2022;12:a040501.
7. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–25.
8. Bowden J, Davey Smith G, Haycock P, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14.
9. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46:1985–98.
10. Burgess S, Thompson S. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181:251–60.
11. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single sample and two-sample summary data settings. Int J Epidemiol. 2019;48:713–27.
12. Sanderson E, Spiller W, Bowden J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. Stat Med. 2021;40:5434–52.
13. Holmes MV, Lange LA, Palmer T, Lanktree MB, North KE, Almoguera B, et al. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. Am J Hum Genet. 2014;94:198–208.
14. Ahmad OS, Morris JA, Mujammami M, Forgetta V, Leong A, Li R, et al. A Mendelian randomization study of the effect of type-2 diabetes on coronary heart disease. Nat Commun. 2015;6:7060.
15. Larsson SC, Mason AM, Bäck M, Klarin D, Damrauer SM, Million Veteran Program. et al. Genetic predisposition to smoking in relation to 14 cardiovascular diseases. Eur Heart J. 2020;14:3304–10.
16. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet. 2018;50:524–37.
17. Bos MJ, Koudstaal PJ, Hofman A, Ikram MA. Modifiable etiological factors and the burden of stroke from the Rotterdam study: a population-based cohort study. PLoS Med. 2014;11:e1001634.
18. Petruski-Ivleva N, Viera AJ, Shimbo D, Muntner P, Avery CL, Schneider AL, et al. Longitudinal patterns of change in systolic blood pressure and incidence of cardiovascular disease: the atherosclerosis risk in communities study. Hypertension. 2016;67:1150–6.
19. Ayala Solares JR, Canoy D, Raimondi FED, Zhu Y, Hassaine A, Salimi-Khorshidi G, et al. Long-term exposure to elevated systolic blood pressure in predicting incident cardiovascular disease: evidence from large-scale routine electronic health records. J Am Heart Assoc. 2019;8:e012129.
20. Cho NH, Cho YR, Park MK, Kim DK, Shin C, Lee MK, et al. Effect of blood pressure on cardiovascular diseases at 10-year follow-up. Am J Cardiol. 2019;123:1654–9.
21. Ghoreishian H, Tohidi M, Derakhshan A, Hajsheikholeslami F, Azizi F, Kazempour-Ardebili S, et al. Presence of hypertension modifies the impact of insulin resistance on incident cardiovascular disease in a Middle Eastern population: the Tehran Lipid and Glucose Study. Diabet Med. 2015;32:1311–8.
22. Redon J, Tellez-Plaza M, Orozco-Beltran D, Gil-Guillen V, Fernandez SP, Navarro-Pérez J, et al. Impact of hypertension on mortality and cardiovascular disease burden in patients with cardiovascular risk factors from a general practice setting: the ESCARVAL-risk study. J Hypertens. 2016;34:1075–83.
23. Twig G, Yaniv G, Levine H, Leiba A, Goldberger N, Derazne E, et al. Body-mass index in 2.3 million adolescents and cardiovascular death in adulthood. N Engl J Med. 2016;374:2430–40.
24. Chen ZhengMing CZ, Iona A, Parish S, Chen YiPing CY, Guo Yu GY, Bragg F, et al. Adiposity and risk of ischaemic and haemorrhagic stroke in 0.5 million Chinese men and women: a prospective cohort study. Lancet Glob Health. 2018;6:e630–40.
25. Li Y, Yatsuya H, Iso H, Yamagishi K, Saito I, Kokubo Y, et al. Body mass index and risks of incident ischemic stroke subtypes: the Japan public health center-based prospective (JPHC) study. J Epidemiol. 2019;29:325–33.
26. Clair C, Rigotti NA, Porneala B, Fox CS, D'Agostino RB, Pencina MJ, et al. Association of smoking cessation and weight change with cardiovascular disease among adults with and without diabetes. JAMA. 2013;309:1014–21.
27. Markidan J, Cole JW, Cronin CA, Merino JG, Phipps MS, Wozniak MA, et al. Smoking and risk of ischemic stroke in young men. Stroke. 2018;49:1276–8.
28. Rosoff DB, Davey Smith G, Mehta N, Clarke TK, Lohoff FW. Evaluating the relationship between alcohol consumption, tobacco use, and cardiovascular disease: a multivariable Mendelian randomization study. PLoS Med. 2020;17:e1003410.
29. Millwood IY, Walters RG, Mei XW, Guo Y, Yang L, Bian Z, et al. Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500,000 men and women in China. Lancet. 2019;393:1831–42.
30. Kubota Y, Iso H, Yamagishi K, Sawada N, Tsugane S, Tamakoshi A, et al. Daily total physical activity and incident stroke: The Japan public health center-based prospective study. Stroke. 2017;48:1730–36.
31. Porter AK, Schilsky S, Evenson KR, Stevens J, Robinson WR, Schreiner PJ, et al. The association of sport and exercise activities with cardiovascular disease risk: the atherosclerosis risk in communities (ARIC) study. J Phys Act Health. 2019;16:698–705.
32. Zhang Y, Tuomilehto J, Jousilahti P, Wang Y, Antikainen R, Hu G, et al. Total and high-density lipoprotein cholesterol and stroke risk. Stroke. 2012;43:1768–74.
33. Hindy G, Engström G, Larsson SC, Traylor M, Markus HS, Dichgans M, et al. Role of blood lipids in the development of ischemic stroke and its subtypes: a Mendelian randomization study. Stroke. 2018;49:820–27.
34. Sun L, Clarke R, Bennett D, Guo Y, Lewington S, Collins R, et al. Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults. Nat Med. 2019;25:569–74.
35. Paynter NP, Balasubramanian R, Giulianini F, Wang DD, Tinker LF, Greenland P, et al. Metabolic predictors of incident coronary heart disease in women. Circulation. 2018;137:841–53.
36. Cavus E, Karakas M, Ojeda F, Kontto J, Sans S, Ferrieres J, et al. Association of circulating metabolites with risk of coronary heart disease in a European population: results from the biomarkers for cardiovascular risk assessment in Europe (BiomarCaRE) consortium. JAMA Cardiol. 2019;4:1270–79.
37. Wang C, Zhang W, Wang Y, Li W, Zhang J, Liu L, et al. Novel associations between sex hormones and diabetic vascular complications in men and postmenopausal women: a cross-sectional study. Cardiovasc Diabetol. 2019;18:97.
38. Madsen TE, Luo X, Huang M, Charytan DM, Ho PM, Wassertheil-Smoller S, et al. Circulating SHBG (sex hormone-binding globulin) and risk of ischemic stroke: findings from the WHI. Stroke. 2020;51:1257–64.

39. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.
40. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9.
41. Clarke TK, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112,117). Mol Psychiatry. 2017;22:1376–84.
42. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47:1121–30.
43. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial. TOAST. Trial of org 10172 in acute stroke treatment. Stroke. 1993;24:35–41.
44. Scholder S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Genetic markers as instrumental variables. J Health Econ. 2016;45:131–48.
45. Bowden J, Spiller W, Del Greco MF, Sheehan N, Thompson J, Minelli C, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. Int J Epidemiol. 2018;47:1264–78.
46. Sanderson E, Windmeijer F. A weak instrument F-test in linear IV models with multiple endogenous variables. J Econom. 2016;2:212–21.
47. Neale Lab. GWAS results from UK Biobank. 2017. Available from: http://www.nealelab.is/uk-biobank/. Accessed January 5, 2023.
48. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408.
49. Kim H, Choe SA, Lee SA, Sung J. Causal relationship between the timing of menarche and young adult body mass index with consideration to a trend of consistently decreasing age at menarche. PLoS ONE. 2021;16:e0247757.
50. Georgakis MK, Gill D, Webb A, Lambertsen KL, Elliott P, Markus HS, et al. Genetically determined blood pressure, antihypertensive drug classes, and risk of stroke subtypes. Neurology. 2020;95:e353–61.
51. Neal B, MacMahon S, Chapman N, Chen Z, Chalmers J, Rodgers A, et al. Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomised trials. Lancet. 2000;356:1955–64.
52. Kaur N, Pandey AK, Negi H, Bhatt SP, Chandra S, Mahdi AA, et al. Effect of HDL-raising drugs on cardiovascular outcomes: a systematic review and meta-regression. PLoS ONE. 2014;9:e94585.
53. McTaggart F, Jones P. Effects of statins on high-density lipoproteins: a potential contribution to cardiovascular benefit. Cardiovasc Drugs Ther. 2008;22:321–38.
54. Sterne JA, Davey Smith G. Sifting the evidence – what's wrong with significance tests? Another comment on the role of statistical methods. BMJ. 2001;322:226.
55. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538:161–64.
56. Hindorff LA, Bonham VL, Brody LC, Ginoza ME, Pan Y, Green ED, et al. Prioritizing diversity in human genomics research. Nat Rev Genet. 2018;19:175–85.
57. Morris AP. Transethnic meta-analysis of genomewide association studies. Genet Epidemiol. 2011;35:809–22.

ACKNOWLEDGEMENTS

The authors would like to thank participants of the UK Biobank from which individual-level data was sourced. Analyses were performed under application number 81499.

AUTHOR CONTRIBUTIONS

YJ conceived the study, performed the data analysis, and drafted the manuscript. WS contributed to statistical analysis and interpretation, and assisted in manuscript writing. ES, KT, and TP provided expert advice on epidemiological and statistical methodology. EH and YK contributed clinical expertise and interpretation of the findings. All authors reviewed and approved the final version of the manuscript.

FUNDING

This work was funded by The Wellcome Trust (108902/B/15/Z) and UK Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC UU 00011/2, MC UU 00011/1). In addition, this research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI21C1243), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT; MSIT) (No. RS-2023-00210888).

COMPETING INTERESTS

The authors declare no competing interests.

ETHICAL APPROVAL

All methods were carried out in accordance with relevant guidelines and regulations. The study was reviewed and approved by the Ewha Womans University Seoul Hospital Institutional Review Board (IRB File No: SEUMC 2021-08-026). The study used de-identified secondary data, and therefore informed consent from individual participants was not required. No identifiable images or personal information of human participants were used in this study.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41371-025-01048-y.

Correspondence and requests for materials should be addressed to Yongho Jee.

Reprints and permission information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025