Mendelian Randomization as a Tool for Cardiovascular Research: A Review

Michael G. Levin, MD[1]; Stephen Burgess, PhD[3,4]

[1]Division of Cardiovascular Medicine, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA USA

[2]MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK

[3]Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Word Count: 3850

Abstract:

Importance: Mendelian randomization (MR) is a statistical approach that has become increasingly popular in the field of cardiovascular disease research. It offers a way to infer causal relationships between risk factors and outcomes using observational data, which is particularly important in cases where randomized controlled trials are not feasible or ethical. With the growing availability of large genetic datasets, MR has become a powerful and accessible tool for studying the risk factors for cardiovascular disease.

Observations: MR leverages genetic variation as a natural experiment to overcome the potential biases that affect traditional observational study designs. It uses genetic variants that are randomly assigned at conception as proxies for exposure to a risk factor, effectively mimicking a randomized controlled trial. By comparing the outcomes of individuals with different genetic variants, researchers can draw causal inferences about the effects of specific risk factors on cardiovascular disease.

Conclusions and Relevance: In this review, we provide an overview of MR methodology, including its underlying assumptions, strengths, and limitations. We also highlight several important applications of MR in cardiovascular disease research, including the identification of novel drug targets, the evaluation of potential cardiovascular risk factors, and potential applications of emerging methodology. Overall, MR offers a powerful approach for investigating causal relationships in observational data and has the potential to transform our understanding of the etiology and treatment of cardiovascular disease.

Introduction:

Understanding the causal relationships between risk factors / biomarkers and health outcomes is key to inform treatment paradigms and therapeutic development. Traditionally, the strongest evidence to support causal relationships comes from randomized controlled trials (RCTs). By randomly allocating participants to treatment and control groups, causal effects can be isolated from competing risk factors that confound traditional observational study designs. While RCTs reliably estimate causal effects, they are often large and costly.[1,2]

Mendelian randomization (MR) offers the opportunity to investigate causal effects from observational data. MR is a natural experiment, analogous to an RCT, which under certain assumptions can overcome the biases that limit traditional observational study designs. MR relies on genetic association studies, which link DNA sequence variants with a variety of risk factors, biomarkers, and health outcomes. Within a population, segregation and independent assortment of DNA sequence variants (according to Mendel's laws) leads to some individuals

having genetically-influenced higher- or lower- levels of a risk factor or exposure of interest, analogous to allocation to treatment and control groups in an RCT (**FIGURE 1**). For example, an RCT to study the effect of LDL-cholesterol (LDL-c) on coronary artery disease (CAD) might randomize individuals to receive either an LDL-lowering treatment or a placebo, and then monitor for the development of CAD. An analogous MR study would compare rates of CAD among individuals with genetically increased or decreased levels of LDL-c.

The proliferation of publicly available genetic association studies has made the MR study design particularly attractive for rapidly evaluating relationships between risk factors and health outcomes. In this review, we provide an overview of the MR methodology and underlying assumptions, discuss strengths and limitations of MR studies, and highlight emerging applications of MR to better understand the mechanisms of cardiovascular disease.

Mendelian Randomization as a Method

Mendelian randomization is typically implemented as a form of instrumental variable (IV) analysis, a statistical method to infer causal relationships that has been developed over the past 100 years.[3] While the relationship between two variables (say LDL-c and CAD) can be estimated using traditional observational study designs (eg. case-control or cohort studies), several factors limit the robustness of these analyses. First, observational studies may be affected by unmeasured confounding, where unmeasured variables associated with both the exposure and outcome lead to biased estimates of the effect of the exposure on the outcome. Second, they may be affected by reverse causality, where the outcome influences the exposure (rather than vice-versa). For example, individuals with CAD may have lower-than-average LDL-c due to the prescription of statins for secondary prevention of cardiovascular events. To overcome these challenges, instrumental variables can divide the population into subgroups, balancing confounding variables, analogous to the randomization step in RCTs.

Instrumental variables must satisfy certain assumptions in order to permit causal inference (**FIGURE 2**):

- **IV1.** The IV must be associated with the exposure (population subgroups have different average levels of the exposure).
- **IV2.** The IV must not associate with the outcome through a confounding pathway (population subgroups have similar levels of competing risk factors).
- **IV3.** The IV must not directly influence the outcome (the instrument may only influence the outcome via the exposure).

MR is a special case of instrumental variable analysis where genetic variants are used as IVs. Genetic variants are particularly attractive as IVs for a few reasons. First, during meiosis, genetic variants are randomly assorted into gametes, and ultimately offspring. This random assortment during meiosis ensures that genetic variants are approximately randomly distributed throughout the population. Second, this random assortment occurs before the onset of disease (protecting against bias due to reverse-causation). When appropriately selected, genetic variants meeting the IV assumptions can be used to enable causal inferences in observational datasets.

Data Sources and Study Design

Table 1 provides an overview of important considerations when performing MR studies. Most MR studies either follow a single-sample or a two-sample design, reflecting whether the genetic associations with the exposure and outcome are assessed in the same (single-sample) or different (two-sample) datasets. Two-sample MR analyses are typically performed without the use of individual-level genotype or phenotype data, leveraging the growing public availability of summary genetic data through resources like the GWAS Catalog, Open GWAS Project, and PhenoScanner.[4–6] These public resources provide access to thousands of genetic association studies across a wide spectrum of traits and diseases, including many relevant to cardiovascular disease. These resources are continuously growing and include genetic associations with many cardiovascular diagnostic traits (eg. electrocardiograms, echocardiograms, and cardiac MRI), and molecular phenotypes (eg. gene expression, metabolite, and protein levels), in addition to typical case/control studies of cardiovascular outcomes.[7] These rich datasets enable researchers to study exposure-outcome pairs that may otherwise be infeasible to consider using local resources. Two-sample approaches can also be implemented using individual-level data; for example, by constructing a genetic score using weights from one sample, and assessing its associations with the outcome in a second sample. Many user-friendly software packages have been developed to perform MR analyses and assess the MR assumptions using summary genetic data.[5,6,8]

Assessing MR Assumptions

When selecting genetic variants as IVs, instrument strength is an important characteristic to consider, as the power of MR studies is influenced by the strength of association between the IV and exposure of interest. Selecting weak instruments (which are not strongly associated with the exposure of interest) can lead to biased estimates of the exposure-outcome relationship.

Genetic variants used as IVs are often selected from large genome-wide association studies (GWAS). Instrument strength can be formally evaluated as the F-statistic from regression of the exposure on the instrument, or the proportion of variance in the exposure explained by the instrument. Beyond GWAS, genetic variants used as IVs may also be identified by other approaches. For example, genetic variants with well-characterized effects on a biological pathway of interest (eg. loss-of-function or other coding variants) or located at specific locations in the genome (eg. in/near a protein-coding gene or drug-target) may represent useful IVs, even if not meeting the stringent significance thresholds commonly employed by large GWAS. For example, common coding variants in the IL6-receptor, which decrease macrophage inflammatory signaling and reduce circulating levels of IL6 and inflammatory markers, have been used to proxy the effects of IL6-inhibitors like tocilizumab, sarilumab, and ziltevekimab. MR studies using these coding variants as genetic instruments have demonstrated the potential for anti-IL6 therapy as a treatment for coronary artery disease, peripheral artery disease, and abdominal aortic aneurysm.[9–12] These MR results predated findings from the CANTOS and RESCUE trials, and laid the groundwork for the ongoing phase III ZEUS trial evaluating ziltivekimab vs. placebo for the prevention of major cardiovascular events among individuals with cardiovascular disease, chronic kidney disease, and inflammation (NCT05021835).

Unlike the first MR assumption, the second (instrument is independent from confounders) and third assumptions (instrument affects the outcome only through the exposure) are more challenging to justify. Genetic variants may have pleiotropic effects (that is, a single variant may affect several traits/diseases via distinct causal pathways), making it difficult to determine with certainty whether a genetic variant influences an outcome through the exposure of interest, or via confounders or direct effects that may violate IV2 and IV3. Selecting genetic variants with well-characterized biological mechanisms may be helpful in justifying these assumptions. Several pleiotropy-robust MR methods have also been developed to detect and/or account for the possibility of pleiotropy, and applying these approaches as sensitivity analyses is key to claiming causal inferences when the IV assumptions cannot be biologically justified. Common robust MR methods include the weighted-median, MR-Egger, and MR-PRESSO methods.[13–15] Dozens of other robust MR methods have been developed, recently reviewed by Sanderson et al.[16] While these methods relax some of the core IV assumptions, they bring additional assumptions which must also be considered. Other approaches to evaluate the plausibility of IV2 and IV3 include the use of negative-control outcomes (eg. hair color), where the exposure-negative control relationship is known to be implausible[17], or negative-control study

populations, such as where the exposure is absent in the population (for example, genetic associations of alcohol-related variants would not be expected in a population group that did not drink alcohol[18]). Unexpected associations with negative controls may indicate violations of the IV assumptions.

Limitations of MR and Sources of Bias

MR can be a powerful tool for causal inference, but like all observational study designs has limitations. Beyond violations of the core MR assumptions, several additional sources of bias may potentially impact findings (**TABLE 2**). While not exhaustive, these sources of bias highlight important considerations for authors and readers when performing and interpreting MR studies.

Reporting and Interpreting MR Results

Guidelines for reporting and interpreting MR studies have been developed by expert consensus, and the Strengthening the Reporting of Observational Studies in Epidemiology initiative provides MR-specific guidance for authors (STROBE-MR: strobe-mr.org).[19,20] In general, authors are encouraged to explicitly consider the underlying assumptions of MR, provide justification for the instruments they select, and cautiously interpret findings. A helpful framework for reporting MR findings is to separate the reporting of results from subsequent causal inferences.[21] First, methods and results are reported without causal language, objectively reporting any associations identified of genetically-predicted exposure levels with outcomes. Second, if appropriate, authors may choose to discuss potential causal interpretations of these associations, justifying the strengths and weakness of those inferences in the context of the core MR assumptions and other relevant literature. Results from MR investigations can rarely prove a causal relationship beyond reasonable doubt, but they may motivate the design of a clinical trial or functional experiment to provide stronger evidential support for a causal relationship.

Although MR conceptually mimics an RCT, it is important to remember several important differences between these study designs when reporting results. Because genetic effects act from conception onwards, MR estimates typically represent lifelong effects, in contrast to the relatively short-term effects typically evaluated in RCTs. Genetic effects also tend to be small in magnitude, and affect the usual/physiologic levels of a trait. In contrast, pharmacologic interventions studied in RCTs may have much larger effects, which occur over a much shorter time-scale (eg. hours/days/weeks) Genetic associations are often transformed into non-clinical

units for statistical efficiency. Estimates from MR and RCTs may not be directly comparable in magnitude, although should correspond in their direction of effects.

Mendelian Randomization Studies in Cardiovascular Disease

Genetic studies of cardiovascular traits were among the first published more than 15 years ago.[22] Since that time, the public availability of large genetic association datasets has dramatically lowered the barrier to performing two-sample MR studies.[23] Because two-sample designs do not require the exposure and outcome to be measured in the same dataset, this approach has allowed investigators to investigate a diverse array of potential exposure-outcome relationships, which may otherwise be infeasible to consider using other study designs. Here, we highlight several scenarios relevant to cardiovascular disease that can be addressed by MR.

Randomized Clinical Trials are Infeasible

Although RCTs represent an important "gold-standard" for estimating causal effects, not all questions are amenable to the RCT design. In some scenarios, randomizing individuals to an exposure may be unethical. For example, while robust epidemiologic evidence has linked smoking to increased rates of atherosclerosis and other cardiovascular diseases, randomizing individuals to a long-term smoking group in a cardiovascular outcomes trial would be unethical. MR has been used to provide additional certainty to conventional observational studies, confirming strong associations between smoking and atherosclerosis.[24] MR has similarly been used to interrogate the cardiovascular consequences of alcohol intake.[25] In other scenarios, RCTs may be too large, expensive, or long in duration in order to be feasible. For example, an RCT to investigate the role of LDL-cholesterol-lowering in young/healthy individuals with low-short term risk of atherosclerosis would require a large sample (tens-of-thousands of participants) to be followed for a long timeframe to identify meaningful effects on cardiovascular outcomes. However, MR findings suggest that lifetime lowering of LDL-cholesterol has strong protective effects on coronary artery disease and overall mortality.[26]

Drug-target Validation

An entire sub-field of MR has developed with a focus on guiding drug discovery and development, known as *cis*-MR, with specific methodological approaches and considerations for assessing the therapeutic potential and adverse effects of drug targets.[27–29] These efforts

have been aided by the growing availability of public datasets cataloging genetic associations with thousands of circulating proteins, metabolites, and expression of genes involved in the regulation of these processes.[30–35] This approach was recently used to demonstrate that genetically-elevated levels of factor XI, a coagulation factor involved in thrombus propagation and stabilization, associate with substantially increased risk of ischemic stroke.[36] Combined with other genetic evidence that suggested factor XI deficiency was associated with reduced rates of cardiovascular events and venous thromboembolism, factor XI inhibition has become an attractive therapeutic target.[37] A recent phase II randomized clinical trial (PACIFIC-AF) investigated the safety of the factor XIa inhibitor asundexian compared with apixaban among individuals with atrial fibrillation, finding asundexian was associated with lower rates of bleeding.[38] The phase III OCEANIC-AF study (NCT05643573) will test the effectiveness of asundexian in preventing stroke and embolic complications among individuals with atrial fibrillation. MR findings demonstrating the protective effects of reducing circulating lipoprotein(a) on atherosclerosis have similarly motivated the development of pharmacologic therapies, which are now in late-stage trials.[39] Several additional targets relevant to cardiometabolic disease are supported by MR evidence, with phase II and III clinical trials underway.[27]

Prioritizing Candidate Biomarkers

Traditional epidemiologic studies have identified hundreds of potential biomarkers for cardiovascular disease, but determining whether these are primarily useful 1) for risk prediction or 2) as therapeutic targets can be challenging to disentangle. Although increased levels of a biomarker may predict poor prognosis, whether those increased levels are the cause or consequence of disease has important implications for the therapeutic potential of the biomarker. For example, a recent study investigated the role of 90 circulating cardiovascular proteins as predictors of heart failure.[40] While 49 proteins were strongly associated with increased heart failure events in an observational analysis, only 8 had evidence of causal relationships by MR, and for 5/8 proteins the MR effects were in a discordant direction from the epidemiologic findings.[40] Similarly, in a landmark MR study, Voight et al. found that HDL-cholesterol, a longstanding predictor of CAD in epidemiologic studies, was unlikely to have causal effects, consistent with RCTs.[41] MR has also been used to clarify the cardiovascular role of uric acid, a biomarker with controversial predictive and causal evidence in the epidemiologic literature.[42] Gill et al. provided evidence that genetically-elevated serum urate increases cardiovascular disease across the coronary, cerebral, and peripheral vascular beds, an effect

likely mediated by increases in systolic blood pressure.[43] Analyses like these, which disentangle the predictive and causal role of biomarkers can aid future target development. Biomarkers without strong evidence of causal effects may be prioritized for biomarker assays and incorporation into predictive models, while biomarkers prioritized by MR may be investigated further for therapeutic development.

Expanding Therapeutic Indications

When designing large-scale cardiovascular outcome trials, pharmaceutical companies may include a large number of outcomes in order to extract the most information from otherwise large and expensive trials. However, including multiple endpoints increases the possibility of false-positive outcomes. The US Food and Drug Administration provides guidance for balancing these considerations, and the entire spectrum of important endpoints may not be included in large cardiovascular outcome studies.[44] For example, although atherosclerotic cardiovascular disease manifests across several vascular beds (eg. coronary arteries, carotid arteries, lower extremity/peripheral arteries, and abdominal aorta, among others), large RCTs often focus on myocardial infarction and stroke as primary endpoints. While RCTs have demonstrated strong treatment effects for lipid- and blood-pressure lowering medications on these endpoints, the RCT evidence for these therapies in other forms of atherosclerosis like peripheral artery disease (PAD) and abdominal aortic aneurysm (AAA) is much less robust.[45–47] When outcomes are biologically and genetically similar, MR may be useful to rapidly evaluate expanded therapeutic indications, without the need for additional large expensive RCTs. Indeed, MR has demonstrated that lipid- and blood-pressure lowering are likely to benefit understudied diseases like PAD and AAA, although the magnitude of benefit may vary across vascular beds.[48–50]

Identifying Adverse Effects

Wide-angled MR approaches have been used to screen for therapeutic effects and on-target toxicities across a broad range of outcomes. This approach can be useful for prioritizing drug targets that are likely to have a favorable risk/benefit profile. For example, in a phenome-wide MR analysis, Georgakis et al. identified beneficial effects of IL6-inhibition across a range of cardiometabolic traits/diseases, while also identifying increased risk of cellulitis and urinary tract infections, highlighting potential adverse effects related to the immunosuppressive mechanism of this target.[9] Similarly, despite potential beneficial effects on cardiovascular outcomes, MR has identified potential adverse effects associated with CETP inhibition,

including increased risk of macular degeneration.[50] Using MR to evaluate the effects of exposures on lifespan can also be a helpful approach to examine the potential overall balance of therapeutic effects and toxicities. For example, Daghlas and Gill demonstrated that increased LDL-cholesterol reduces lifespan, and that LDL-reduction by PCSK9 inhibition would be expected to significantly improve longevity, suggesting an overall net-benefit of this therapeutic class on mortality.[26,52]

Future Directions for Mendelian Randomization in Cardiovascular Disease Research

Although MR studies investigating cardiovascular risk factors and outcomes have been pursued for nearly two decades, new datasets and methodologic developments are poised to enable ongoing insights. Here, we discuss several active areas of MR research with potential to improve our understanding of cardiovascular disease.

Interactions and Non-linear Relationships

Epidemiologic studies often test for interactions between risk factors, which may identify combinations of risk factors that disproportionately increase risk of disease, or risk factors that have heterogenous effects across population subgroups. Because non-linear associations in epidemiologic models may be particularly susceptible to confounding, MR approaches can be useful for more accurately modelling potential causal effects. Several MR methods have been developed to interrogate these relationships, including non-linear and factorial MR approaches, which have been applied across the spectrum of cardiovascular disease.[53,54] Non-linear MR can be useful for delineating complex relationships between risk factors and disease outcomes. For example, using data from UK Biobank, Sulc et al. found evidence that anthropometric measures of obesity including body mass index (BMI), waist-hip ratio (WHR), body fat percentage, and others had non-linear effects on many cardiometabolic traits including measures of cholesterol, blood pressure, and glucose.[55] These results highlight the complex relationships between body composition and cardiometabolic health. Arvanitis et al. used non-linear MR to investigate reports from the epidemiologic literature that had suggested diastolic blood pressure may have a J-shaped effect on risk of myocardial infarction.[56] In contrast to epidemiologic findings, the authors found no evidence for a non-linear relationship with myocardial infarction, other cardiovascular outcomes, or mortality, suggesting that blood pressure reduction is likely to have constant beneficial effects, consistent with findings from RCTs. Similarly, Biddinger et al. recently applied non-linear MR to demonstrate that there is

likely no protective level of alcohol intake, in contrast to epidemiologic findings that suggested low-levels of alcohol intake may have cardiovascular benefits.[25]

However, concerns have been raised about the validity of results from non-linear MR methods. Some results from a non-linear MR study investigating the impact of vitamin D were logically impossible, with discrepancies between overall results and stratum-specific estimates[57]. Further methodological investigations have shown that the residual stratification method commonly used for non-linear MR is sensitive to an assumption that the genetic effect on the exposure is constant in the population, and can lead to highly biased results when this assumption is violated[58]. A recent method, the doubly-ranked method, does not make this assumption and so should give less biased results[59]. The reliability of results from these methods is a topic of active research.

Factorial MR can be used to evaluate for interactions between risk factors, identifying population subgroups at particularly high or low risk of disease. For example, Ference et al. applied factorial MR to investigate the relationship between circulating LDL-cholesterol, triglycerides, and coronary heart disease.[60] The authors found that individuals with both genetically-elevated LDL and triglycerides were at higher risk of coronary heart disease than individuals with either risk factor in isolation (or neither risk factor), motivating opportunities to targeting lipoproteins beyond LDL-cholesterol to reduce risk of atherosclerosis. Overall, both non-linear and factorial MR approaches can be especially useful for confirming or refuting epidemiologic findings and identifying population subgroups that may be particularly susceptible to disease.

Time-varying Exposures and Disease Progression

Risk factors for cardiovascular disease (eg. obesity) vary at different stages of the life course (eg. childhood, adolescence, adulthood). Despite this potential variability in the timing of risk factor onset, MR estimates most commonly leverage genetic datasets reflecting adulthood exposures. Recent extensions of multivariable MR have been applied to evaluate the role of risk factors across the life-course. For example, using this approach authors found that childhood body size influences cardiovascular disease indirectly through effects on adult body size, while in contrast childhood body size more directly influences MRI measures of adult cardiac structure.[61,62] A methodological challenge of this approach is that the multivariable MR model treats measures of the risk factor at different timepoints as independent causal factors. This

may be reasonable if the measurements reflect distinct biological mechanisms. But if the measurements at different timepoints simply reflect the same risk factor measured at different arbitrary timepoints, rather than biologically distinct risk factors (as they are for childhood versus adulthood BMI), then the approach is less reliable.[63]

While GWAS have typically focused on snapshots of adult-onset traits and diseases at a single point in time, life-course MR approaches will become increasingly possible as genetic studies of cardiovascular traits and diseases among younger individuals become available.[64] As GWAS of disease severity become available, MR can be a useful tool for studying factors associated with disease progression over time, overcoming index-event/collider bias that may affect traditional observational studies.[65] Finally, the increasing use of wearable devices may allow for the study of new time-varying exposures that have been previously unmeasured at population-scale, including continuous measures of physical activity, vital signs (eg. heart rate, blood pressure, respiration, pulse oximetry), and measures of cardiac structure/function, among others.[66,67]

Disease Mechanisms

Molecular datasets provide the opportunity to better characterize the pathways leading from genetic variant to outcome, delineating the intermediate cells, tissues, and pathways relevant to the disease of interest. Characterizing the genetic regulation of tissue- and cell-specific molecular phenotypes (gene expression, protein abundance, metabolites) is a major focus of genomic research, with important conceptual and methodologic applications to MR.[33,68] Conceptually, MR studies leveraging tissue- and cell-specific genetic instruments may be useful to rapidly define the specific tissue- and cell-types where a drug target is likely to have strong therapeutic effects while minimizing toxicities. Methodologically, several MR methods have been developed which seek to cluster genetic instruments based on similarity of their genetic effects.[69–70] These clustering approaches offer the theoretical potential to identify groups of genetic variants which influence the outcome through a common mechanism, helping to better characterize genetic instruments which may capture several underlying biological processes. These prioritization approaches may ultimately aid downstream development of pharmacologic agents acting on specific pathways, tissues, and cell-types, potentially shortening the timeline from target discovery to human studies.

Role of Ancestry

The vast majority of GWAS and MR studies have been performed among populations of predominately European genetic ancestry.[71] As cardiovascular disease represents the major source of morbidity and mortality worldwide, robustly characterizing the role of risk factors and biomarkers across genetic backgrounds using MR is an important goal. Although recent studies suggest that causal genetic effects are likely to be similar across genetic backgrounds, evaluating the generalizability of MR findings across genetic ancestries has been challenging due to differences in linkage disequilibrium and identification of ancestry-specific causal genetic variants.[72] Ongoing efforts to improve the diversity of participants in GWAS, including the All of US and Veterans Affairs Million Veteran Program in the United States, and the Genes & Health program in the UK, will provide the opportunity to robustly evaluate risk factor-disease relationships across a broader range of genetic backgrounds.[73,74]

Conclusion:

Mendelian randomization is a powerful study design which can overcome important limitations of traditional observational studies. While MR studies require a set of assumptions, when interpreted in the context of other available evidence, MR studies can increase (or decrease) confidence in epidemiological associations and provide strong evidence to motivate rigorous mechanistic and cardiovascular outcome studies.

References:

1. Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA*. 2020;323(9):844-853. doi:10.1001/jama.2020.1166
2. Moore TJ, Zhang H, Anderson G, Alexander GC. Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015-2016. *JAMA Intern Med*. 2018;178(11):1451-1457. doi:10.1001/jamainternmed.2018.3931
3. Stock JH, Trebbi F. Retrospectives: Who Invented Instrumental Variable Regression? *Journal of Economic Perspectives*. 2003;17(3):177-194. doi:10.1257/089533003769204416
4. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:e34408. doi:10.7554/eLife.34408
5. Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(D1):D977-D985. doi:10.1093/nar/gkac1010
6. Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics (Oxford, England)*. 2019;35(22):4851-4853. doi:10.1093/bioinformatics/btz469
7. Kerimov N, Hayhurst JD, Peikova K, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet*. 2021;53(9):1290-1299. doi:10.1038/s41588-021-00924-w
8. Elsworth B, Lyon M, Alexander T, et al. The MRC IEU OpenGWAS data infrastructure. Published online August 10, 2020:2020.08.10.244293. doi:10.1101/2020.08.10.244293
9. Georgakis MK, Malik R, Li X, et al. Genetically Downregulated Interleukin-6 Signaling Is Associated With a Favorable Cardiometabolic Profile: A Phenome-Wide Association Study. *Circulation*. 2021;143(11):1177-1180. doi:10.1161/CIRCULATIONAHA.120.052604
10. Levin MG, Klarin D, Georgakis MK, et al. A Missense Variant in the IL-6 Receptor and Protection From Peripheral Artery Disease. *Circ Res*. 2021;129(10):968-970. doi:10.1161/CIRCRESAHA.121.319589
11. Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium, Swerdlow DI, Holmes MV, et al. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *The Lancet*. 2012;379(9822):1214-1224. doi:10.1016/S0140-6736(12)60110-X
12. Sarwar N, Butterworth AS, Freitag DF, et al. Interleukin-6 receptor pathways in coronary heart disease: A collaborative meta-analysis of 82 studies. *The Lancet*. 2012;379(9822):1205-1213. doi:10.1016/S0140-6736(11)61931-4

13. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic epidemiology*. 2016;40(4):304-314. doi:10.1002/gepi.21965
14. Verbanck M, Chen CY, Neale B, Do R. Widespread pleiotropy confounds causal relationships between complex traits and diseases inferred from Mendelian randomization. *Nature genetics*. Published online 2017:157552. doi:10.1101/157552
15. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology*. 2017;32(5):377-389. doi:10.1007/s10654-017-0255-x
16. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. *Nat Rev Methods Primers*. 2022;2(1):1-21. doi:10.1038/s43586-021-00092-5
17. Sanderson E, Richardson TG, Hemani G, Davey Smith G. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. International Journal of Epidemiology. 2021;50(4):1350-1361.
18. Cho Y, Shin S-Y, Won S, Relton CL, Davey Smith G, Shin M-J. Alcohol intake and cardiovascular risk factors: A Mendelian randomisation study. Scientific Reports. 2015/12/21 2015;5(1):18422.
19. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res*. 2020;4:186. doi:10.12688/wellcomeopenres.15555.2
20. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ*. 2021;375:n2233. doi:10.1136/bmj.n2233
21. Burgess S, O'Donnell CJ, Gill D. Expressing Results From a Mendelian Randomization Analysis: Separating Results From Inferences. *JAMA Cardiology*. 2021;6(1):7-8. doi:10.1001/jamacardio.2020.4317
22. Burton PR, Clayton DG, Cardon LR, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-678. doi:10.1038/nature05911
23. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC- InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015;30(7):543-552. doi:10.1007/s10654-015-0011-z

24. Levin MG, Klarin D, Assimes TL, et al. Genetics of Smoking and Risk of Atherosclerotic Cardiovascular Diseases: A Mendelian Randomization Study. *JAMA Netw Open*. 2021;4(1):e2034461. doi:10.1001/jamanetworkopen.2020.34461
25. Biddinger KJ, Emdin CA, Haas ME, et al. Association of Habitual Alcohol Intake With Risk of Cardiovascular Disease. *JAMA Network Open*. 2022;5(3):e223849. doi:10.1001/jamanetworkopen.2022.3849
26. Daghlas I, Gill D. Low-density lipoprotein cholesterol and lifespan: A Mendelian randomization study. *British Journal of Clinical Pharmacology*. 2021;87(10):3916-3924. doi:10.1111/bcp.14811
27. Burgess S, Mason AM, Grant AJ, et al. Using genetic association data to guide drug discovery and development: Review of methods and applications. *Am J Hum Genet*. 2023;110(2):195-214. doi:10.1016/j.ajhg.2022.12.017
28. Gill D, Georgakis MK, Walker VM, et al. Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res*. 2021;6. doi:10.12688/wellcomeopenres.16544.2
29. Schmidt AF, Finan C, Gordillo-Marañón M, et al. Genetic drug target validation using Mendelian randomisation. *Nature Communications*. 2020;11(1):3255. doi:10.1038/s41467-020-16969-0
30. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet*. 2021;53(12):1712-1721. doi:10.1038/s41588-021-00978-w
31. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558(7708):73-79. doi:10.1038/s41586-018-0175-2
32. Borges MC, Haycock PC, Zheng J, et al. Role of circulating polyunsaturated fatty acids on cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid genetic association data from over 114,000 UK Biobank participants. *BMC Med*. 2022;20(1):210. doi:10.1186/s12916-022-02399-w
33. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-1330. doi:10.1126/science.aaz1776
34. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Mapping the proteo-genomic convergence of human diseases. *Science*. 2021;374(6569):eabj1541. doi:10.1126/science.abj1541

35. Gaziano L, Giambartolomei C, Pereira AC, et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat Med*. 2021;27(4):668-676. doi:10.1038/s41591-021-01310-z

36. Gill D, Georgakis MK, Laffan M, et al. Genetically Determined FXI (Factor XI) Levels and Risk of Stroke. *Stroke*. 2018;49(11):2761-2763. doi:10.1161/STROKEAHA.118.022792

37. Preis M, Hirsch J, Kotler A, et al. Factor XI deficiency is associated with lower risk for cardiovascular and venous thromboembolism events. *Blood*. 2017;129(9):1210-1215. doi:10.1182/blood-2016-09-742262

38. Piccini JP, Caso V, Connolly SJ, et al. Safety of the oral factor XIa inhibitor asundexian compared with apixaban in patients with atrial fibrillation (PACIFIC-AF): a multicentre, randomised, double-blind, double-dummy, dose-finding phase 2 study. *The Lancet*. 2022;399(10333):1383-1390. doi:10.1016/S0140-6736(22)00456-1

39. Burgess S, Ference BA, Staley JR, et al. Association of LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization Analysis. *JAMA Cardiol*. 2018;3(7):619-627. doi:10.1001/jamacardio.2018.1470

40. Henry A, Gordillo-Marañón M, Finan C, et al. Therapeutic Targets for Heart Failure Identified Using Proteomics and Mendelian Randomization. *Circulation*. 2022;145(16):1205-1217. doi:10.1161/CIRCULATIONAHA.121.056663

41. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *The Lancet*. 2012;380(9841):572-580. doi:10.1016/S0140-6736(12)60312-2

42. Feig DI, Kang DH, Johnson RJ. Uric Acid and Cardiovascular Risk. *New England Journal of Medicine*. 2008;359(17):1811-1821. doi:10.1056/NEJMra0800885

43. Gill D, Cameron AC, Burgess S, et al. Urate, Blood Pressure, and Cardiovascular Disease. *Hypertension*. 2021;77(2):383-392. doi:10.1161/HYPERTENSIONAHA.120.16547

44. Research C for DE and. Multiple Endpoints in Clinical Trials Guidance for Industry. U.S. Food and Drug Administration. Published October 20, 2022. Accessed March 21, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry

45. Aboyans V, Ricco JB, Bartelink MLEL, et al. 2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS): Document covering atherosclerotic disease of extracranial carotid

and vertebral, mesenteric, renal,. *European Heart Journal*. 2017;39(9):763-816. doi:10.1093/eurheartj/ehx095

46. Gerhard-Herman MD, Gornik HL, Barrett C, et al. 2016 Lower Extremity Peripheral Arterial Disease Guidelines. *Journal of the American College of Cardiology*. 2017;69(11):e71-e126. doi:10.1016/j.jacc.2016.11.007

47. Isselbacher EM, Preventza O, Hamilton Black J, et al. 2022 ACC/AHA Guideline for the Diagnosis and Management of Aortic Disease: A Report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. *Circulation*. 2022;146(24):e334-e482. doi:10.1161/CIR.0000000000001106

48. Levin MG, Zuber V, Walker VM, et al. Prioritizing the Role of Major Lipoproteins and Subfractions as Risk Factors for Peripheral Artery Disease. *Circulation*. 2021;144(5):353-364. doi:10.1161/CIRCULATIONAHA.121.053797

49. Levin MG, Klarin D, Walker VM, et al. Association Between Genetic Variation in Blood Pressure and Increased Lifetime Risk of Peripheral Artery Disease. *Arterioscler Thromb Vasc Biol*. 2021;41(6):2027-2034. doi:10.1161/ATVBAHA.120.315482

50. Roychowdhury T, Klarin D, Levin MG, et al. Multi-ancestry GWAS deciphers genetic architecture of abdominal aortic aneurysm and highlights PCSK9 as a therapeutic target. Published online May 29, 2022:2022.05.27.22275607. doi:10.1101/2022.05.27.22275607

51. Nordestgaard LT, Christoffersen M, Lauridsen BK, et al. Long-term Benefits and Harms Associated With Genetic Cholesteryl Ester Transfer Protein Deficiency in the General Population. *JAMA Cardiology*. Published online October 6, 2021. doi:10.1001/jamacardio.2021.3728

52. Ference BA, Kastelein JJP, Ray KK, et al. Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants with Risk of Coronary Heart Disease. *JAMA*. 2019;321(4):364-373. doi:10.1001/jama.2018.20045

53. Ference BA, Kastelein JJP, Ginsberg HN, et al. Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk. *JAMA*. 2017;318(10):947-956. doi:10.1001/jama.2017.11467

54. Staley JR, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genetic Epidemiology*. 2017;41(4):341-352. doi:10.1002/gepi.22041

55. Sulc J, Sjaarda J, Kutalik Z. Polynomial Mendelian randomization reveals non-linear causal effects for obesity-related traits. *Human Genetics and Genomics Advances*. 2022;3(3):100124. doi:10.1016/j.xhgg.2022.100124

56. Arvanitis M, Qi G, Bhatt DL, et al. Linear and Nonlinear Mendelian Randomization Analyses of the Association Between Diastolic Blood Pressure and Cardiovascular Events. *Circulation*. 2021;143(9):895-906. doi:10.1161/CIRCULATIONAHA.120.049819

57. Davey Smith G. Mendelian randomisation and vitamin D: the importance of model assumptions. The Lancet Diabetes and Endocrinology. 2023;11(1):P14.

58. Burgess S. Violation of the constant genetic effect assumption can result in biased estimates for non-linear Mendelian randomization. medRxiv. 2022:2022.10.26.22280570. doi:10.1101/2022.10.26.22280570

59. Tian H, Mason AM, Liu C, Burgess S. Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method. PLOS Genetics. 2023;19(6):e1010823. doi:10.1371/journal.pgen.1010823

60. O'Nunain K, Park C, Urquijo H, et al. A lifecourse mendelian randomization study highlights the long-term influence of childhood body size on later life heart structure. *PLoS Biol*. 2022;20(6):e3001656. doi:10.1371/journal.pbio.3001656

61. Power GM, Tyrrell J, Frayling TM, Davey Smith G, Richardson TG. Mendelian Randomization Analyses Suggest Childhood Body Size Indirectly Influences End Points From Across the Cardiovascular Disease Spectrum Through Adult Body Size. *Journal of the American Heart Association*. 2021;10(17):e021503. doi:10.1161/JAHA.121.021503

62. Tian H, Burgess S. Estimation of time-varying causal effects with multivariable Mendelian randomization: some cautionary notes. *International Journal of Epidemiology*. Published online January 20, 2023:dyac240. doi:10.1093/ije/dyac240

63. Richardson TG, O'Nunain K, Relton CL, Davey Smith G. Harnessing Whole Genome Polygenic Risk Scores to Stratify Individuals Based on Cardiometabolic Risk Factors and Biomarkers at Age 10 in the Lifecourse—Brief Report. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2022;42(3):362-365. doi:10.1161/ATVBAHA.121.316650

64. Mitchell RE, Hartley AE, Walker VM, et al. Strategies to investigate and mitigate collider bias in genetic and Mendelian randomisation studies of disease progression. *PLOS Genetics*. 2023;19(2):e1010596. doi:10.1371/journal.pgen.1010596

65. Hu H, Huang H, Li M, et al. A wearable cardiac ultrasound imager. *Nature*. 2023;613(7945):667-675. doi:10.1038/s41586-022-05498-z

66. Dashti HS, Jones SE, Wood AR, et al. Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat Commun*. 2019;10(1):1100. doi:10.1038/s41467-019-08917-4

67. Donovan MKR, D'Antonio-Chronowska A, D'Antonio M, Frazer KA. Cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. *Nat Commun*. 2020;11(1):955. doi:10.1038/s41467-020-14561-0

68. Foley CN, Mason AM, Kirk PDW, Burgess S. MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates. *Bioinformatics*. 2021;37(4):531-541. doi:10.1093/bioinformatics/btaa778

69. Grant AJ, Gill D, Kirk PDW, Burgess S. Noise-augmented directional clustering of genetic association data identifies distinct mechanisms underlying obesity. *PLOS Genetics*. 2022;18(1):e1009975. doi:10.1371/journal.pgen.1009975

70. Liang X, Mounier N, Apfel N, et al. MR-AHC: A fast, efficient and robust method for two-sample summary data Mendelian randomization based on agglomerative hierarchical clustering. Published online March 23, 2023. doi:10.1101/2023.03.18.23287164

71. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell*. 2019;177(1):26-31. doi:10.1016/j.cell.2019.02.048

72. Hou K, Ding Y, Xu Z, et al. Causal effects on complex traits are similar for common variants across segments of different continental ancestries within admixed individuals. *Nat Genet*. Published online March 20, 2023:1-10. doi:10.1038/s41588-023-01338-6

73. The "All of Us" Research Program. *New England Journal of Medicine*. 2019;381(7):668-676. doi:10.1056/NEJMsr1809937

74. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology*. 2016;70:214-223. doi:10.1016/j.jclinepi.2015.09.016

75. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Statistical methods in medical research*. 2017;26(5):2333-2355. doi:10.1177/0962280215597579

76. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genetic Epidemiology*. 2016;40(7):597-608. doi:10.1002/gepi.21998

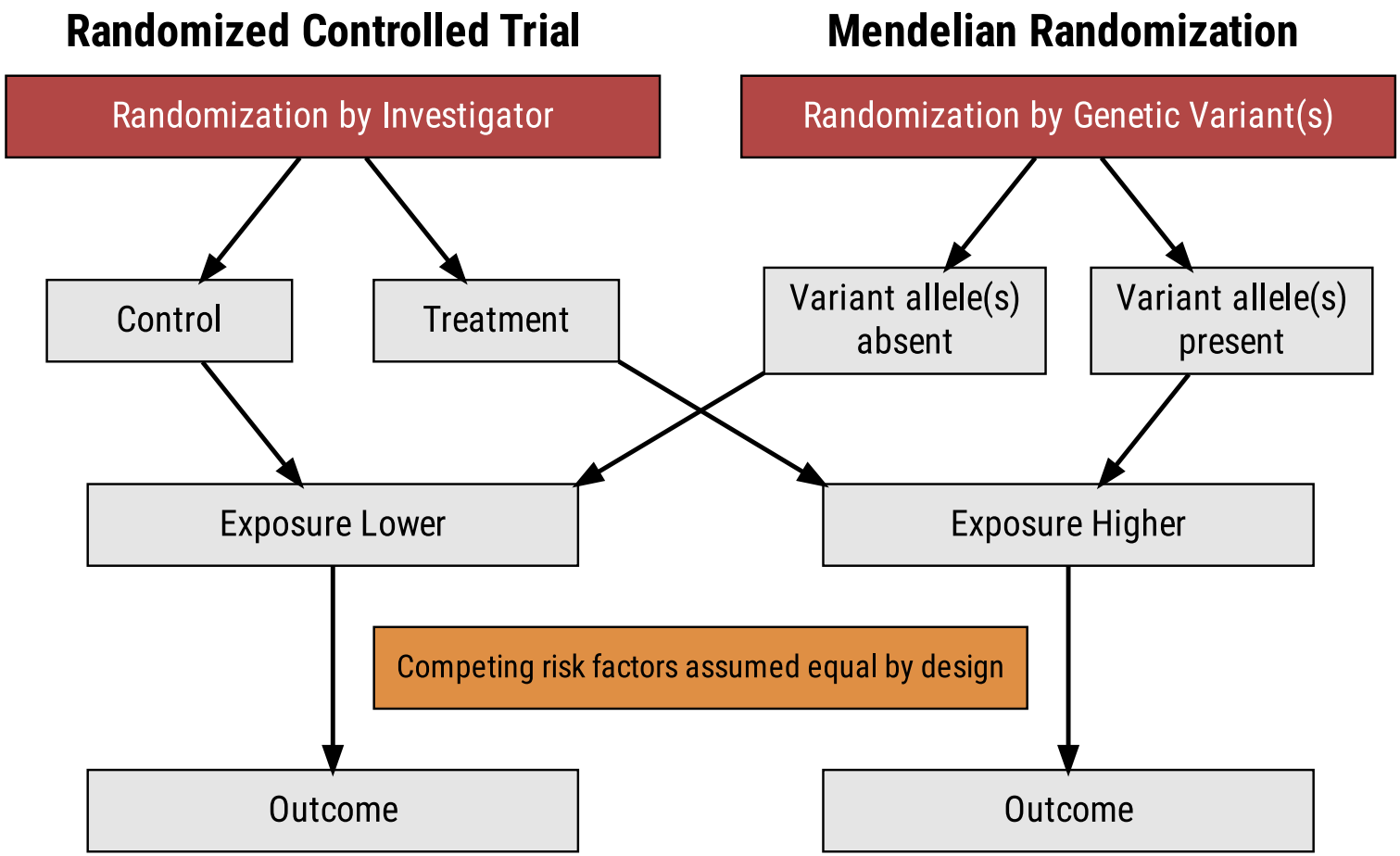
77. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27(8):1133-1163. doi:10.1002/sim.3034

78. Fatumo S, Karhunen V, Chikowore T, et al. Metabolic Traits and Stroke Risk in Individuals of African Ancestry: Mendelian Randomization Analysis. *Stroke*. 2021;52(8):2680-2684. doi:10.1161/STROKEAHA.121.034747

79. Jiang T, Gill D, Butterworth AS, Burgess S. An empirical investigation into the impact of winner's curse on estimates from Mendelian randomization. *International Journal of Epidemiology*. Published online December 27, 2022:dyac233. doi:10.1093/ije/dyac233

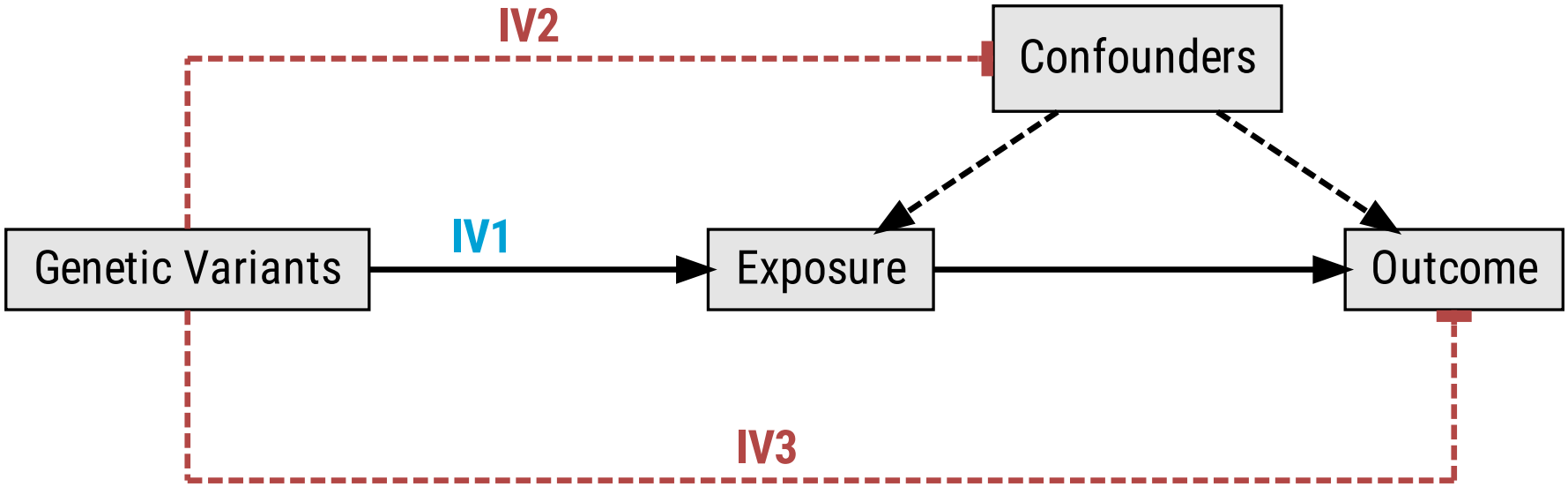
80. Walker V, Harrison S, Carter A, Gill D, Tzoulaki I, Davies N. The consequences of adjustment, correction and selection in genome-wide association studies used for two-sample Mendelian randomization [version 1; peer review: 1 approved with reservations]. *Wellcome Open Research*. 2021;6(103). doi:10.12688/wellcomeopenres.16752.1

Figure 1 – MR as a Naturally Randomized Trial



Comparison of Randomized Controlled Trial (RCT) and Mendelian Randomization (MR) study designs. In an RCT, the eligible population is randomized to treatment or control groups. In MR studies, the population is divided into subgroups based on the presence or absence of genetic variants, randomly allocated at conception, that are associated with the exposure of interest. In both cases, the population becomes divided into groups with different average levels of the exposure, and randomization ensures that competing risk factors are equally balanced between groups. The outcome of interest can be measured in each group and differences between groups can be evaluated using statistical tests.

Figure 2 – MR + Instrumental Variable Assumptions



Directed acyclic graph of an MR study, representing the relationships between the instrumental variable, exposure, outcome, and potential confounders. MR depends on core instrumental variable assumptions in order to justify causal inferences. IV1: The IV must be associated with the exposure (population subgroups have different average levels of the exposure). IV2: The IV must not associate with the outcome through a confounding pathway (population subgroups have similar levels of competing risk factors). IV3: The IV must not directly influence the outcome (the instrument may only influence the outcome via the exposure).

TABLE 1 – MR Study Design Considerations

Characteristic	Considerations
Study question	MR studies are designed to support causal inference, rather than risk-prediction. Common questions that may be suitable for MR include: • Does a particular risk factor cause a specific disease of interest? • Is a specific lifestyle factor, such as physical activity or diet, causally related to the incidence of a certain disease? • Is a particular drug target likely to prevent or treat a specific disease? • Is a circulating protein or metabolite likely to cause a certain disease?
Data sources	In one-sample designs, genotype, exposure, and outcome data are obtained from a single study. In two-sample designs, exposure and outcome data (usually summary data on genetic associations) are obtained from two or more studies. The OpenGWAS Project, PhenoScanner, and GWAS Catalogue are useful public resources for obtaining summary genetic data for two-sample studies.
Instrumental variable(s)	Single or multiple genetic variants can be used as instrumental variables, but must be strongly associated with the risk factor/exposure of interest and satisfy the other MR assumptions. Well-characterized biological function can often support the use of single genetic variants. Multiple genetic variants, identified from large genetic association studies, are often used to improve statistical power, but may introduce pleiotropic effects.
MR Methods	MR methods make different assumptions. It is common for MR studies to utilize several methods in order to evaluate the robustness of the results. In individual-level designs, the ratio of coefficients or two-stage least squares methods are common primary analysis methods. In summary data designs, the ratio of coefficients and inverse variance weighted methods are common primary analysis methods.
Diagnostics and Sensitivity analyses	Both graphical and statistical approaches can be used to evaluate some of the MR assumptions. Scatter-, funnel-, and forest-plots of individual genetic variants can help identify potential outliers, which may be pleiotropic variants. Bi-directional MR can be used to assess whether the outcome influences the exposure, which may indicate genuine bi-directional causal effects or a common genetic architecture. Genetic colocalization can be used to support MR findings focused on a specific region of the genome (eg. a specific drug target or protein-coding gene).

Interpretation of Results	MR results should be interpreted cautiously, as they can rarely prove a causal relationship beyond reasonable doubt. Causal claims should be justified based on the plausibility of the MR assumptions and in the context of other forms of evidence.
Statistical software	Many statistical programming packages have been developed to perform MR analyses. Common packages for the R programming language include the *OneSampleMR* package for single-sample studies, and the *TwoSampleMR* and *MendelianRandomization* packages for two-sample study designs.

TABLE 2 – Sources of Bias in MR Studies

Bias	Description
Weak Instruments	Weak instruments, which are not strongly associated with the exposure of interest, can lead to biased MR estimates. In the single-sample setting, this bias occurs toward the observational estimate; while in the two-sample setting, this bias is toward a null association.[70] If the exposure and outcome samples have common participants, then bias is between these two cases depending on the degree of sample overlap.[71]
Linkage Disequilibrium	Genetic instruments may be correlated with variants that are associated with the outcome of interest due to linkage disequilibrium (LD), violating the MR assumptions.[72] To minimize bias due to differences in patterns of allele frequency or LD between ancestry groups, MR studies are typically stratified by genetic ancestry. Because the majority of genetic association studies to-date have been performed on individuals of European ancestry, most MR studies have consequently focused on this population.[66] While non-European populations remain under-represented in MR, a recent multi-ancestry MR study investigated the role of diabetes and circulating lipids as risk factors for ischemic stroke, finding similar effects among African- and European-ancestry populations.[73] Efforts to increase the representativeness and generalizability of MR findings to diverse populations are critically needed and ongoing.
Winner's Curse	Winner's curse refers to the tendency to overestimate genetic associations in the GWAS that first detected the variant. Winner's curse leads to biased association estimates, particularly for genetic instruments just reaching the genome-wide level for statistical significance; although in an empirical example, the impact of this bias on MR estimates was modest.[74] The robustness of MR findings to winner's curse may be assessed by restricting the analysis to genetic instruments most-strongly associated with the exposure of interest.
Selection and Survival Bias	Although genetic variants are randomly allocated at conception, GWAS typically enroll adult populations. Therefore, differential selection of study participants, or differences in survival between birth and enrollment could lead to biased association

	estimates. However, simulation studies indicate that selection effects need to be fairly substantial in magnitude to adversely affect conclusions from MR studies.[75]