

Causal inference in health and disease: a review of the principles and applications of Mendelian randomization

Catherine E. Lovegrove^{1,*} , Sarah A. Howles¹ , Dominic Furniss², Michael V. Holmes³

¹Nuffield Department of Surgical Sciences, University of Oxford, Oxford OX3 9DU, United Kingdom

²Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, United Kingdom

³Medical Research Council, Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, United Kingdom

*Corresponding author: Catherine E. Lovegrove, Nuffield Department of Surgical Sciences, University of Oxford, Oxford OX3 9DU, United Kingdom (catherine.lovegrove@nds.ox.ac.uk).

Abstract

Mendelian randomization (MR) is a genetic epidemiological technique that uses genetic variation to infer causal relationships between modifiable exposures and outcome variables. Conventional observational epidemiological studies are subject to bias from a range of sources; MR analyses can offer an advantage in that they are less prone to bias as they use genetic variants inherited at conception as “instrumental variables,” which are proxies of an exposure. However, as with all research tools, MR studies must be carefully designed to yield valuable insights into causal relationships between exposures and outcomes, and to avoid biased or misleading results that undermine the validity of the causal inferences drawn from the study. In this review, we outline Mendel’s laws of inheritance, the assumptions and principles that underlie MR, MR study designs and methods, and how MR analyses can be applied and reported. Using the example of serum phosphate concentrations on liability to kidney stone disease we illustrate how MR estimates may be visualized and, finally, we contextualize MR in bone and mineral research including exemplifying how this technique could be employed to inform clinical studies and future guidelines concerning BMD and fracture risk. This review provides a framework to enhance understanding of how MR may be used to triangulate evidence and progress research in bone and mineral metabolism as we strive to infer causal effects in health and disease.

Keywords: epidemiology, genetic association studies, human association studies, mendelian randomization, therapeutics

Lay Summary

Mendelian randomization is a powerful tool used by researchers to understand how factors like lifestyle choices or medical conditions affect our health. Some research studies can be hampered by inaccurate information or measurements which mean the results are unreliable. Mendelian randomization uses genetic information to evade this problem. Because a person’s genetic information does not change the studies can provide more reliable conclusions about how different factors influence our health. This review provides an overview of Mendelian randomization principles, its applications, and how researchers in bone and mineral research have used it to uncover cause-and-effect relationships.

Introduction

Defining disease-causing pathways is of central importance in medical research and can facilitate interventions to improve health. Conventional observational epidemiological approaches can allude to associations between exposures and outcomes, but are subject to limitations including the inability to detect the direction of causal relationships, bias from measurement error, or residual confounding from external factors that are not fully accounted for (Figure 1).^{1,2} These limitations may lead to erroneous conclusions regarding the nature of the relationship between an exposure and outcome. Well-designed randomized control trials (RCTs) are widely considered the “gold standard” approach to overcoming these problems to ascertain causal relationships,³ however not all research questions are suited to an RCT, for example where the outcome of interest is rare or takes a long time to manifest. Even when a hypothesis is amenable to testing in an RCT, there are practical barriers to conducting such a study. Furthermore, in cases where an exposure is known to be harmful, such as

obesity, an RCT that randomizes a group of individuals to interventions aiming to establish a BMI over 30 kg/m² would be unethical.^{4,5}

Mendelian randomization (MR) is an epidemiological technique that aims to overcome the limitations of conventional epidemiological studies, and the practical difficulties associated with an RCT, by leveraging genetic variation to identify causal effects between an exposure and outcome.⁶ An analogy can be drawn between MR and an RCT in that MR exploits genetic alleles to allocate individuals to a “study arm” (Figure 2).⁷ The analogy has some limitations and cannot fully capture the intricacies of identifying causal effects, however it can be a helpful framework help conceptualize MR. Using genetic alleles, inherited at conception, to produce an unbiased proxy of an exposure trait, an “instrumental variable” (IV, Table 1), helps to mitigate against possible effects from environmental confounders (Table 1).^{8–11} Researchers commonly identify genetic variants associated with exposures and outcomes to use in MR studies from genome-wide association

Received: March 19, 2024. Revised: July 4, 2024. Accepted: August 19, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the American Society for Bone and Mineral Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Terminology in MR studies.

Term	Definition
Confounder	A variable associated with the exposure and the outcome that can induce a biased relationship between an exposure and outcome. This can be potentially accounted for in multivariable analyses; however, if overlooked, it can result in unreliable results
Horizontal pleiotropy	A genetic variant affects the outcome through traits independent of the hypothesized exposure
Instrumental variable	A genetic proxy for an exposure variable comprising genetic variants that are robustly associated with that variable
Linkage disequilibrium	A non-random distribution of genetic variants in the population due to their physical proximity on chromosomes causing them to be inherited together
Mediator	An intermediate variable through which an exposure affects an outcome
One-sample MR	An MR analysis that derives an IV exposure from the same dataset in which it conducts the MR analysis with the outcome
Regression dilution	Random errors in measurement of the exposure leading to bias toward the null in the estimated association of an exposure with an outcome
Residual confounding	Confounding that persists despite adjusting for measured confounders
Reverse causality	An effect of an outcome on an exposure often leading to full, or partial, misattribution of causality in the incorrect direction
Two-sample MR	An MR analysis that uses an exposure IV and an outcome IV derived from 2 separate datasets
Vertical pleiotropy	A genetic variant influences multiple traits (including the outcome variable of interest) through 1 causal pathway
Weak instrument bias	Bias arising from an IV that is weakly associated with the exposure variable, thus violating the “relevance” assumption

Abbreviations: IV, instrumental variable; MR, Mendelian randomization

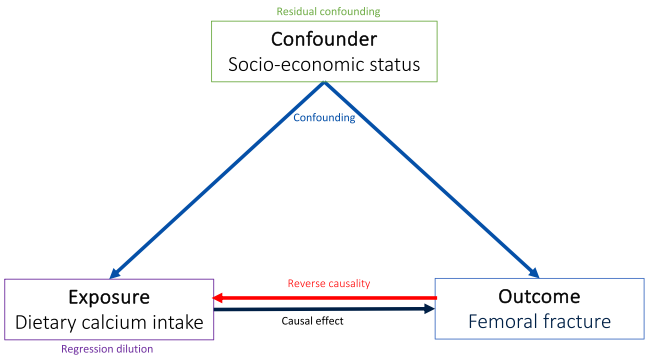


Figure 1. Pitfalls in observational studies. Confounding variable affecting exposure and outcome: measured confounders can be accounted for in multivariable analyses; unmeasured confounders can cause unreliable results. Residual confounding: confounding that persists despite adjusting for measured confounders. Reverse causality: possible effect of outcome on exposure, often neglected. Regression dilution bias: random errors in measurement causing bias in estimated association with outcome.

studies (GWAS); the effect on an outcome per unit increase in a genetically-predicted continuous exposure or the occurrence of an outcome in these genetically-predefined groups is subsequently assessed.

MR studies have become increasingly common likely, in part, due to their potential to elucidate causal relationships and the widespread availability of GWAS data.^{12,13} Despite thousands of MR studies being published annually, many are unreliable as they don’t meet accepted standards.^{14,15} Just as an RCT must be carefully executed to avoid bias, meticulous design of MR studies is essential to ensure that analyses reliably evaluate the presence or absence of causal effects of an exposure on an outcome.¹⁶ Before using MR, researchers should understand its advantages and pitfalls. They ought also to have a clear, relevant, biologically plausible hypothesis with a strong rationale for using this technique. With the burgeoning popularity of MR in the scientific literature, this review aims to provide the readership with the tools to understand, critically appraise, and design MR studies.

Assumptions and principles of MR

Mendel’s laws of genetic inheritance underpin the principles of MR.⁸ The law of “independent segregation” states that at every site on an autosome, 1 allele is inherited from the mother and 1 allele from the father, and which allele inherited from each parent is random.^{9,10} The law of “independent assortment” states that inheritance of alleles is independent of other inherited alleles or traits, except for genomic regions that are inherited together, known as linkage disequilibrium (LD) (Table 1).^{9,10,17} Based on these principles, MR studies that use genetic variants as an IV for an exposure can avoid bias from confounding variables (Table 1). To reliably draw causal inference from MR, 3 conditions must be met, and authors of MR studies should, as far as possible, demonstrate that these assumptions have not been violated (Figure 3):

Condition 1- Relevance: The IV must be reliably associated with the exposure.

Condition 2- Exchangeability or independence: The exposure IV is not caused by a factor that also causes the outcome, that is, no unmeasured confounding variables between the exposure IV and the outcome.

Condition 3- Exclusion restriction: The exposure IV only affects the outcome via the exposure trait and not by any other trait with a downstream effect on the outcome of interest.

Furthermore, robust MR findings rely on an assumption of gene–environment equivalence, that is, a change in an exposure due to genetic variation has the same downstream effects on a phenotype as an environmental or pharmaceutical influence.¹⁸ Not all genetic variation exactly mimics environmental alterations thus gene–environment equivalence must be considered when interpreting the reliability of results from MR.¹⁶ For example, genotypic influences on serum phosphate concentration or a similar increase in serum phosphate concentration influenced by dietary intake or supplementation would be expected to result in the same effects on reducing risk of kidney stone disease.¹⁹ Although there are not yet observational data that corroborate this finding, a meta-analysis showed that increasing phosphate intake lowers urine calcium, which provides preliminary triangulating evidence that higher phosphate might lead to a lower

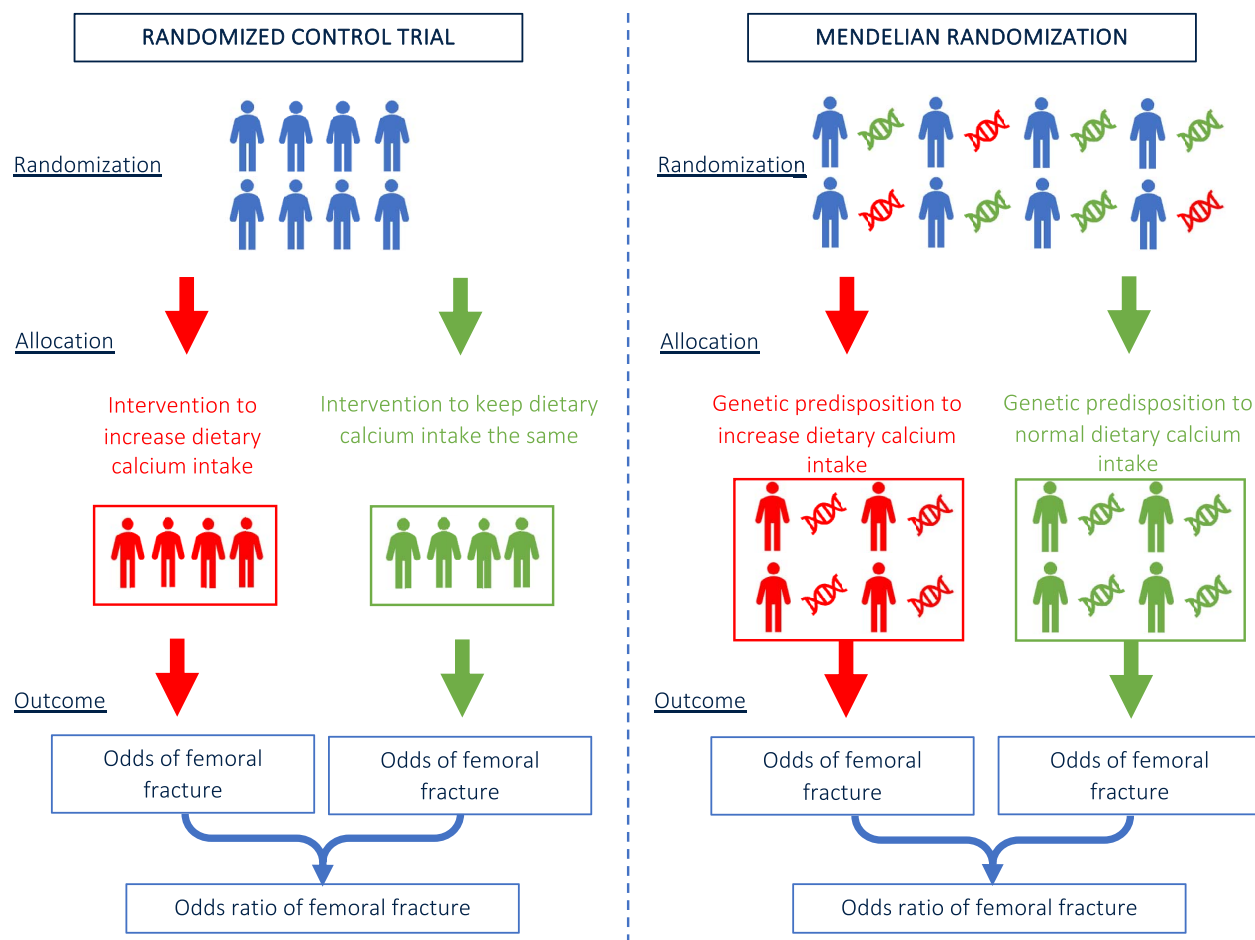


Figure 2. Analogy of an RCT and MR. MR leverages naturally occurring genetic variations as “instrumental variables” to mimic the random assignment of exposures, like random allocation in an RCT. This genetic randomization helps address confounding factors and strengthens causal inference when investigating the relationships between exposures, outcomes, and potential interventions.

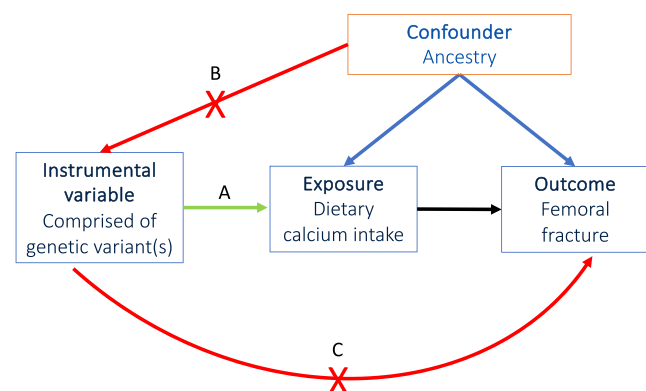


Figure 3. Conditional assumptions in MR. (A) Relevance: The genetic variant(s) acting as a proxy for the exposure variable are strongly associated with the exposure. (B) Independence: The IV is independent of confounding factors, thus genetic association with confounding variables is random, akin to randomization in an RCT. (C) Exclusion restriction: The IV affects the outcome solely through the exposure variable and not by external, horizontally-pleiotropic, pathways.

risk of kidney stones.²⁰ While gene–environment equivalence cannot be “proven,” it is essential that researchers appraise whether it is a reasonable assumption considering available evidence regarding the biological mechanisms underlying the

exposure–outcome relationship. For example, MR of lipid-lowering therapies such as statins (instrumented by genetic variants associated with the gene encoding HMG-CoA reductase) or PCSK9 inhibitors provides empirical evidence of gene–environment equivalence: MR yields genetically predicted causal effects that triangulate with findings from large phase III trials.²¹ A nuance is that the long, subclinical phase of atherosclerosis preceding symptomatic cardiovascular disease means that MR estimates of effect are much larger than those from shorter clinical trials. Understanding the scaling factors for each exposure–disease pair can help scientists to consider what results could be expected from a clinical trial.

Another important consideration in MR studies is the temporality of exposure. Using genetic variants inherited at conception as IVs means that MR study results reflect a lifelong exposure. For example, MR analysis of blood pressure on coronary artery disease represents years of exposure to hypertension rather than the follow-up period of an RCT designed to evaluate the effects of anti-hypertensive medication on coronary artery disease. Moreover, such studies assume that associations of genetic variants with an exposure are constant,²² however effects of an exposure on an outcome may relate to the level of the exposure at a certain stage in life. For example, evidence suggests that early life adiposity influences risk of breast cancer independent of later life adiposity.²³

When using MR, it is important that the IV is derived from the same underlying population in which the outcome is assessed.^{16,24} For example, causal inference would be invalid if a study investigated the effect of alcohol intake using an exposure IV derived from a population of individuals who consume alcohol with an outcome IV derived from a population of individuals with little or no alcohol intake, for example women in some east Asian cultures.²⁵ Considering genetic ancestry is crucial as patterns of LD vary across ancestral groups, hence if different ancestral groups are employed in MR analyses, a genetic variant may not have the same association with an exposure trait in exposure and outcome datasets.²⁶ Reliance on ancestry-specific GWASs, combined with the current Eurocentric focus and lack of well-powered studies in non-European populations, exacerbates historic health research disparities and limits generalizability across populations.²⁷

Further assumptions made in MR studies are that genetic instruments demonstrate linear associations with an outcome and are homogeneous, that is, the relationship between an IV and exposure is the same throughout the studied population.²⁸

MR study design

MR can be undertaken using either individual-level genetic and phenotype data or from GWAS summary statistics. The latter allows researchers to consider larger sample sizes and more easily share data across institutions. As sample size, case-control ratio, and variance explained by IVs increase in a GWAS with a binary endpoint, so does the statistical power and precision of the study.¹⁶ Individual genetic variants may be used as discrete instruments or combined into an allele score generated from the sum of risk alleles.²⁹ Allele scores should be weighted to reflect the estimated effect of each variant on an exposure to account for different variant-effect associations. Combining appropriately-selected variants into an allele score can increase the power of analyses which is particularly useful when individual variants have a small effect on the exposure, represented by the beta-coefficient or odds-ratio, and using individual variants could hypothetically violate the relevance assumption of MR (weak instrument bias, Table 1).²⁹ Using multiple variants in MR is akin to combining data from multiple trials; where multiple variants are used it is important that these are independent (ie, not in LD) to prevent the same association signal being considered twice. Therefore, prior to performing MR, variants in an exposure IV that are in LD should be “clumped” to retain only 1 variant per locus, using a suitable population linkage pattern. Alternatively, an LD matrix can be included in analyses to account for correlation between variants.³⁰

Because common genetic variants typically account for a small proportion in the variation of a trait, large GWAS datasets are frequently used to derive well-powered MR studies. Multiple consortia and GWAS catalogs provide publicly-available summary statistics from GWAS and meta-analyses that facilitate these analyses. [Supplementary Table S1](#) details some large GWAS consortia and catalogs used in MR studies.^{6,31–42} For example, the Genetic Factors for Osteoporosis Consortium (GEFOS) is an international collaboration that used data from 20 439 fracture cases and 78 843 controls to identify risk factors for osteoporotic fractures using summary MR.⁴³

One-sample MR

One-sample MR refers to analyses performed using data from a single population. Estimates derived from 1-sample analysis may be subject to overfitting, that is, bias away from the null hypothesis, leading to an incorrect assumption of causality.^{11,29,44} The risk of overfitting arises because the exposure IV is used to predict the outcome in the same population it was derived from, risking chance correlations with potential confounding variables.⁴⁵ This can be particularly problematic in the presence of weak instrument bias, an issue that can limit both 1- and 2-sample MR studies. However, as MR has advanced and researchers now use stronger instruments from large consortia and databases, the risk of overfitting has decreased.

Two-sample MR

Two-sample MR analyses assess the effect of an exposure on an outcome in a dataset that is different from that in which the exposure IV was derived. This has particular utility when a dataset is limited by small sample size or a lack of appropriate biospecimens; researchers can leverage larger datasets for genetic associations with the exposure and outcome separately, which can improve the statistical power of analysis.¹¹ Two-sample MR assumes that the populations from which the IV associations are derived share equivalent genetic ancestry, age, and sex distributions.^{26,46} It is important that there is adequate case ascertainment in the population from which the outcome IV is derived to ensure that MR estimates are well-powered. Increasing use of GWAS from biobanks and consortia can result in sample overlap in 2-sample MR.⁴⁶ The effect of overlap on a continuous outcome is proportional to the degree of overlap between samples; as overlap increases the study design more closely resembles that of 1-sample MR. For binary outcomes the effects of overlap depend on whether case participants' data have been used to derive exposure IV⁴⁵; if only control participants overlap between exposure and outcome datasets then unbiased estimates are achieved, however if both case and control participants overlap then the effect of bias is comparable to that on a continuous outcome.⁴⁵ Recent analyses propose that sample overlap may not be as problematic as previously suggested. Power, bias, and type 1 error rate in the context of sample overlap can be estimated using tools such as the web application <https://sb452.shinyapps.io/overlap>.⁴⁵

MR estimates

Wald ratio

The simplest method for deriving an estimate for the causal effect of a single IV on an outcome is the Wald ratio (Table 1). This is the ratio of the regression of the outcome on the IV and the regression of the exposure on the IV.⁴⁷ For continuous outcomes, the resulting estimate is typically expressed as the change in outcome per unit change in exposure due to the variant of interest. For binary outcomes, the Wald ratio is interpreted as the log odds ratio for disease risk per unit increase in the exposure due to the variant of interest. A Wald ratio intimates that a variant may exert causal effects on an outcome; however, one must be mindful that the variant may be in LD with an outcome-associated variant, an issue that can also affect an IV comprised of multiple genetic variants, and may lead to erroneous conclusions.

Inverse variance weighted estimates

To obtain an overall MR estimate for an IV comprising multiple variants, researchers often conduct an inverse variance weighted (IVW) meta-analysis of each genetic variant's Wald ratio (Table 1). Conceptually, the IVW model can be considered as a regression of variant-outcome effects against variant-exposure effects weighted according to the inverse of the variance of the variant-outcome effects. The intercept of this regression is fixed at zero (ie, a variant with no effect on the exposure has no effect on the outcome).

A fixed effect IVW is calculated assuming that all variants within the IV uphold the assumptions of MR. When all MR assumptions are met, the IVW estimate has the highest statistical power and is often used alongside other MR methods as sensitivity analyses.⁴⁸ However, should any variants in the IV exhibit horizontal pleiotropy (Table 1), that is, they influence the outcome via pathways other than the exposure of interest, then the IVW estimate may be biased.⁴⁸ Balanced horizontal pleiotropy (the deviation of a variant's estimate from the mean estimate is independent of its effect on the exposure) may be accounted for if the random effects do not systematically affect the outcome. Here, the IVW method can provide an unbiased estimate of causal effect.⁴⁹

MR-Egger estimates

If there is directional horizontal pleiotropy and variants in an IV affect the outcome through pathways not mediated by the exposure (Table 1, arrow C Figure 3), the IVW estimate is biased and the assumption of exchangeability is violated. This can be overcome by enabling a “non-zero” intercept when regressing variant-outcome effects against variant-exposure effects, allowing for directional (ie, “unbalanced”) horizontal pleiotropy (Table 1).^{50,51} In MR-Egger, the intercept estimate represents the average pleiotropic effect across all variants and the slope is an estimate of the causal effect after accounting for pleiotropy.

The MR-Egger estimate can be unbiased even if all variants in an IV exhibit horizontal pleiotropy provided that the effects of horizontal pleiotropy are not related to variant-exposure effects (the InSIDE assumption (Instrument Strength Independent of Direct Effect)). For example, if a study investigates the relationship between stress levels and the risk of coronary heart disease and variants in the exposure IV are associated with smoking, it is plausible that these variants affect risk of heart disease by influencing both smoking and stress levels. If the genetic effects on smoking behaviors are correlated with the effects of genetic variants on stress levels then the InSIDE assumption is violated and the MR-Egger estimate will be biased.⁵² Moreover, MR-Egger estimates are unreliable when there are few genetic variants in an IV because the precision of the estimate decreases as the variability of variant-exposure associations increases.⁵⁰

Weighted median

An alternative method of addressing unbalanced pleiotropy is to use the median effect of all variants within an IV (Table 1).⁵³ Whereas the simple median estimate provides a consistent causal estimate if half of variants in an IV are valid, the weighted median provides a consistent estimate when at least half of the weight comes from valid variants that do not violate the assumptions of MR. Variants are weighted by the inverse variance of their association with the outcome;

thus, variants with stronger associations contribute more to the weighted median estimate.

Simple mode and weighted mode

Mode-based estimates generate clusters of variants, using a kernel density function, with each group having similar causal effects (Table 1). The resulting estimate is derived from the cluster comprising the most variants.⁵⁴ If the largest cluster comprises valid variants, the mode-based estimate will be unbiased. A weighted estimate involves weighting each variant's contribution to its cluster according to the inverse variance of its association with the outcome.

MR pleiotropy residual sum and outlier

The MR-pleiotropy residual sum and outlier (MR-PRESSO) method aims to identify horizontally pleiotropic outliers to overcome bias that arising from MR estimates influenced by horizontal pleiotropy (Table 1).⁵⁵ This method applies 3 sequential tests to summary-level MR: identifying horizontal pleiotropy (global test), correcting pleiotropy by removing outlying variants (outlier test), and ascertaining whether there are significant differences in causal estimates after outlying variants have been removed (distortion test).⁵⁵

Multivariable MR

It may be relevant to undertake multivariable analysis to ascertain the effect of an exposure on an outcome after controlling for relevant covariates (Table 1). Multivariable MR is of particular utility when the “*exclusion restriction*” assumption has been violated, as it allows genetic variants to have pleiotropic effects on the outcome via other risk factors.⁵⁶ Furthermore, multivariable MR allows multiple, correlated, exposures to be included in analyses, and enables researchers to ascertain if all included exposure variables exert direct effects on an outcome and to identify mediating relationships providing that the F-statistic (see “Assessing the Core Assumptions of MR” section) is sufficiently high and thus there are sufficient genetic differences between included traits. Multivariable MR has been used to allude to the role of adiposity and risk of kidney stone disease as risk of kidney stone disease increases with higher BMI, independent of waist-to-hip-ratio (WHR), and with higher WHR independent of BMI.¹⁹

Multivariable MR relies on the same assumptions as univariable MR.⁵⁷ Furthermore, multivariable MR assumes that the included genetic variants affect the outcome only via the risk factors included in analysis.⁵⁶

Mediation MR

Mediation analyses enable exploration of the pathways by which an exposure affects an outcome, and estimation of the proportion of the effect of the exposure on the outcome mediated by an additional trait (Figure 4, Table 1).⁵⁸

Mediation MR analyses can provide an estimate of the effect of an exposure on an outcome via all possible pathways (total effect), an estimate of the effect of an exposure on the outcome via mechanisms independent of the mediating factor(s) (direct effect), and an estimate of the effect of the exposure on the outcome via the specified mediator(s) (indirect effect).⁵⁸ These analyses have been used to demonstrate that higher WHR mediates 12%-15% of its effect on kidney stone risk through increasing serum calcium concentrations,¹⁹ and that higher WHR and BMI increase the risk of coronary

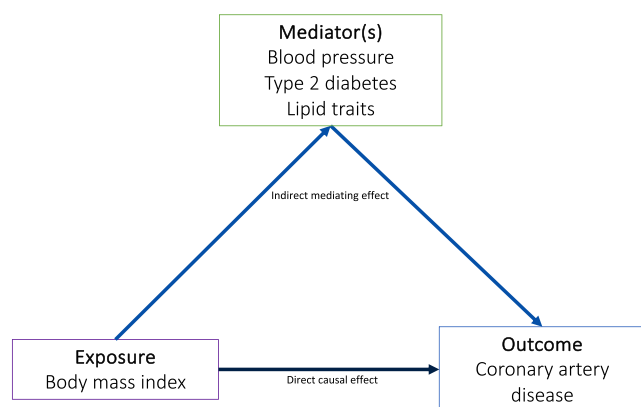


Figure 4. Mediation MR. Mediation analysis estimates the effect of an exposure on an outcome via mediating factors (indirect mediating effect) and via mechanisms independent of the mediating factor (direct causal effect). The proportion mediated effect can subsequently be derived.

artery disease via effects mediated by systolic blood pressure (~27%), type 2 diabetes (~41%), and lipid traits (~3%).⁵⁹

Mediation analysis can be undertaken either using 2-step or multivariable MR methods. Using 2-step MR, the effect of the exposure on the mediator and the effect of the mediator on the outcome can be multiplied to ascertain the indirect effect (product of coefficients method).⁶⁰ In multivariable MR methods, the direct effect of the exposure on the outcome, controlling for the mediator, is estimated. The indirect effect is calculated by subtracting the “controlled” direct effect from the total effect (difference method).⁶⁰ The “proportion mediated effect” can be derived to estimate the proportion of the total effect of an exposure on an outcome explained by a mediating pathway (Figure 4).

Non-linear MR

If associations between an exposure and outcome are more complex than a linear function, MR results from parametric methods, including IVW, MR-Egger, mode-, and median-based estimates, will be unreliable; non-linear MR methods may confirm or refute a causal effect (Table 1),^{61–63} where there is a relationship between an exposure and outcome. Moreover, non-linear MR analyses can identify population subsets that are more vulnerable to disease.

For example, a (now withdrawn) study of serum 25-hydroxyvitamin-D concentrations and cardiovascular outcomes and mortality suggested that increasing vitamin-D concentrations in populations with low vitamin-D concentrations may decrease mortality.⁶⁴ However, subsequent analyses in the same population revealed that the constant genetic effect assumption had been violated.⁶⁵ When the authors reanalyzed data using the non-parametric stratification “doubly ranked” method, which does not rely on the constant genetic effect assumption, no evidence of a link between vitamin-D and mortality at any vitamin-D concentration was identified.⁶¹

Conversely, observational analyses have described a J-shaped relationship between BMI and all-cause mortality.⁶⁶ A causal effect of BMI on all-cause mortality was shown in linear and non-linear MR estimates although the J-shaped curve was flatter in non-linear MR estimates than in conventional, observational, epidemiological analyses. Following meta-regression analyses, authors demonstrated that the non-linear relationship was driven by extreme quantiles of BMI where datapoints were scarce. After removing the first and 99th

percentiles of BMI data MR estimates indicated a linear effect of BMI on mortality.⁶⁶

Recent analyses have discussed problems related to non-linear MR methods.⁶⁷ Wade et al. adopted a negative control methodology to evaluate these methods. Their results that higher BMI increases the odds of being of male sex highlight the biases of residual and doubly-ranked based approaches. In the same candid article the authors express concern over the reliability of results in publications drawing conclusions from non-linear MR analyses including their own “>70 other papers”. Overall, while MR is a decades-old field, there remain complexities in reliably instrumenting polygenic traits.

Drug-target MR

MR has been applied to predict drug effects and toxicities in humans⁶⁸ and provide evidence for the likely utility of a drug prior to embarking on costly clinical trials. Assuming that a drug-target is a protein, MR can be employed to proxy effects of targeting this protein by using genetic variants that are a valid IV of the target protein and estimating the effect of modulating this protein on an outcome. Restricting the IV to variants within a defined distance of transcription start and stop sites (cis protein quantitative loci) increases confidence that identified effects are exerted via the gene encoding the protein.⁶⁹ Many drug-target MR analyses triangulate evidence, including genetic variants associated with a biomarker affected by a drug’s target. Drug-target MR estimates the effect of changing this biomarker, by altering the drug target’s activity, on an outcome. For example, statins and proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors are used to reduce low-density lipoprotein cholesterol (LDL-c) to prevent cardiovascular disease.^{70,71} However, lower LDL-c is associated with a higher incidence of type 2 diabetes.^{72–74} Schmidt et al. identified variants at the PCSK9 locus associated with reductions in LDL-c and demonstrated that lower LDL-c results in higher body weight, WHR, fasting plasma glucose, HbA1c, and liability to type 2 diabetes.⁷⁵ As previously mentioned, the magnitude of effect from a MR study may not mirror that of a clinical intervention as MR estimates represent life-long exposure. Furthermore, causal estimates derived from MR do not necessarily mirror the effects seen in clinical practice, as demonstrated by the hitherto lack of effect of PCSK9 inhibition on increasing the risk of type 2 diabetes in RCTs.^{76–79}

A further use of MR in drug-development is through phenome-wide association studies (PheWAS) to explore a wide range of effects of modulating drug targets.^{80–83} Wang et al. used a genetic proxy of angiopoietin-like proteins 3 and 4 (ANGPTL3 and ANGPTL4) and lipoprotein lipase (LPL) to demonstrate an association with lipoprotein lipids and fatty acids, and employed drug-target MR to reveal that modulating coronary heart disease and type 2 diabetes risk at the ANGPTL4 and LPL loci was predicted to be clinically beneficial with concomitant predicted effects on WHR, whereas an ANGPTL3 variant showed a distinct effect on renal function.⁸⁴ These findings replicated those reported in clinical studies.^{85–88}

Visual representation of MR analyses

There are multiple methods to visualize MR findings, each providing specific information to facilitate assessment of the reliability of a study and its underlying assumptions.¹⁹ A common representation of summary-data MR is a scatter plot

of associations of individual variants with the exposure and outcome (Figure 5a). Error bars, representing the standard error of the estimate in the exposure and outcome, surround each point and regression lines for each sensitivity analysis may be overlaid showing the overall estimate of the analysis. Scatter plots reveal whether variants have similar estimates of effect or if there is heterogeneity and the overall estimate is driven by a single genetic variant (Figure 5a). Individual variant-specific causal effects can be plotted as a forest plot and compared with overall IVW and MR-Egger estimates (Figure 5b). Both scatter and forest plots can help to identify heterogeneity in the causal estimates from different variants.²⁶ Funnel plots enable assessment of asymmetry in causal estimates and directional pleiotropy if estimates from weaker genetic variants are biased in 1 direction (Figure 5c).

Leave-one-out analysis plots represent the effect of removing 1 genetic variant at a time on the overall IVW MR estimate, enabling one to identify a genetic variant with an important influence on the overall result (Figure 5d). Using this information, researchers can assess how robust their findings are to removing these variants.

Radial MR plots enable one to visualize the weight attributed to each variant included in the MR estimate and identify outliers (Figure 5e).⁸⁹ The distance between each point and the IVW slope represents the contribution of the genetic variant to Cochran's Q-statistic of heterogeneity.⁸⁹

Assessing the core assumptions of MR

Whereas combining genetic variants into an IV increases power, there is a risk of "weak instrument bias" because each variant has a small effect and may violate the "relevance" assumption. Weak instruments reduce the power to detect a causal effect and may result in a biased MR estimate.⁴⁵ In 1-sample MR this estimate is biased toward identifying a relationship between the risk factor and outcome, whereas in 2-sample MR it is biased toward the null.⁴⁶ Careful selection of genetic variants for IVs, for example ensuring that variants reach GWAS-significance ($p < 5 \times 10^{-8}$), can reduce bias from a weak instrument. The F-statistic quantifies the strength of an IV by regressing an exposure on a set of variants (Figure 6b).⁴⁷ The higher the F-statistic, the more robust the instrument; a score greater than 10 indicates that the IVs are sufficiently strong for univariable and multivariable MR analyses.⁵⁷ An F-statistic less than 10 does not confirm that an instrument is insufficiently strong but indicates that researchers should consider the possibility of weak instrument bias within their analyses.

If all genetic variants are valid instruments, one would expect the relationship between variants, the exposure, and outcome to be linear and homogenous for all individuals within a population, and for there to be no more heterogeneity in variant-exposure and variant-outcome estimates than explainable by chance alone.²⁶ However, this theory is seldom, if ever, satisfied. Heterogeneity may arise from variants acting on a trait through a variety of physiological mechanisms or from non-linear⁷² relationships between variants, exposures, and outcomes within a population. Heterogeneity can be calculated for IVW and MR-Egger estimates using Cochran's Q or the I^2 statistic which tests for horizontal pleiotropy (Figure 6c).⁵⁴ Quantifying heterogeneity enables researchers to identify outliers and mitigate for horizontal pleiotropy, for

example by employing MR-PRESSO.²⁶ Heterogeneity can be visualized using scatter, forest, funnel, or radial plots reflecting the Wald ratio of each variant within an IV (Figure 5).⁴⁶

Variants should explain more variance in the exposure trait than the outcome and the estimated effect of the IV on the outcome should be wholly mediated through the exposure.²⁶ Measurement error, sample size discrepancies between phenotypes, or a combination of quantitative and binary phenotypes can cause a genetic variant to seemingly explain more variation in a more distal trait than a more proximal trait.⁹⁰ For example, 1,25-hydroxyvitamin-D influences albumin-adjusted serum calcium concentrations; if a genetic variant primarily associated with 1,25-hydroxyvitamin-D is included within an IV for albumin-adjusted serum calcium concentrations then misleading estimates for the causal effect of albumin-adjusted serum calcium concentrations on a range of outcomes may arise. To mitigate against this problem and address possible misspecification of the primary exposure "Steiger filtering" or "MR Steiger" can be performed (Figure 6d).⁹¹ This entails removing genetic variants that explain more variation in the outcome than the exposure and assessing the impact of this process on the MR estimate.⁹⁰ MR Steiger can provide evidence supporting the direction of a causal relationship and performing bidirectional MR (analyzing the effect of the exposure on the outcome, and the effect of the outcome on the exposure) using separate sets of IVs can also assess the direction of causal effects.

Using positive and negative control outcomes can be valuable when assessing how robust an MR analysis is.²⁶ A positive control outcome is one known to be caused by the exposure variable; for example, systolic blood pressure is known to increase risk of stroke.^{92,93} It is logical that genetic variants associated with systolic blood pressure will reveal that higher systolic blood pressure increases risk of stroke on MR analyses; if the analyses failed to identify this association, this may indicate that the genetic variants are not valid instruments to proxy systolic blood pressure. Conversely, a negative control outcome is one not understood to be caused by the exposure variable.^{26,94} Thus, genetic variants associated with coffee consumption may have causal effects on risk of heart disease but would not be expected to affect whether people have blue eyes. In this example, eye color acts as a negative control; if the MR estimate shows a significant relationship between the IV proxying coffee consumption and risk of heart disease but not on eye color it would support the validity of the IV and the integrity of the resulting estimate. Conversely, were the MR estimate to show a relationship between the IV for coffee consumption and blue eye color, one should be suspicious of bias, population stratification, or pleiotropy within the IV.

MR in bone and mineral research

To contextualize the possible application of MR to bone and mineral research for readers, we have identified a small selection of studies in the field:

BMD and osteoarthritis

Observational analyses indicate that high BMD is a risk factor for osteoarthritis.⁹⁵ Hartley et al. used MR to ascertain whether this was a causal relationship, if the effect was unidirectional, and whether the relationship was independent

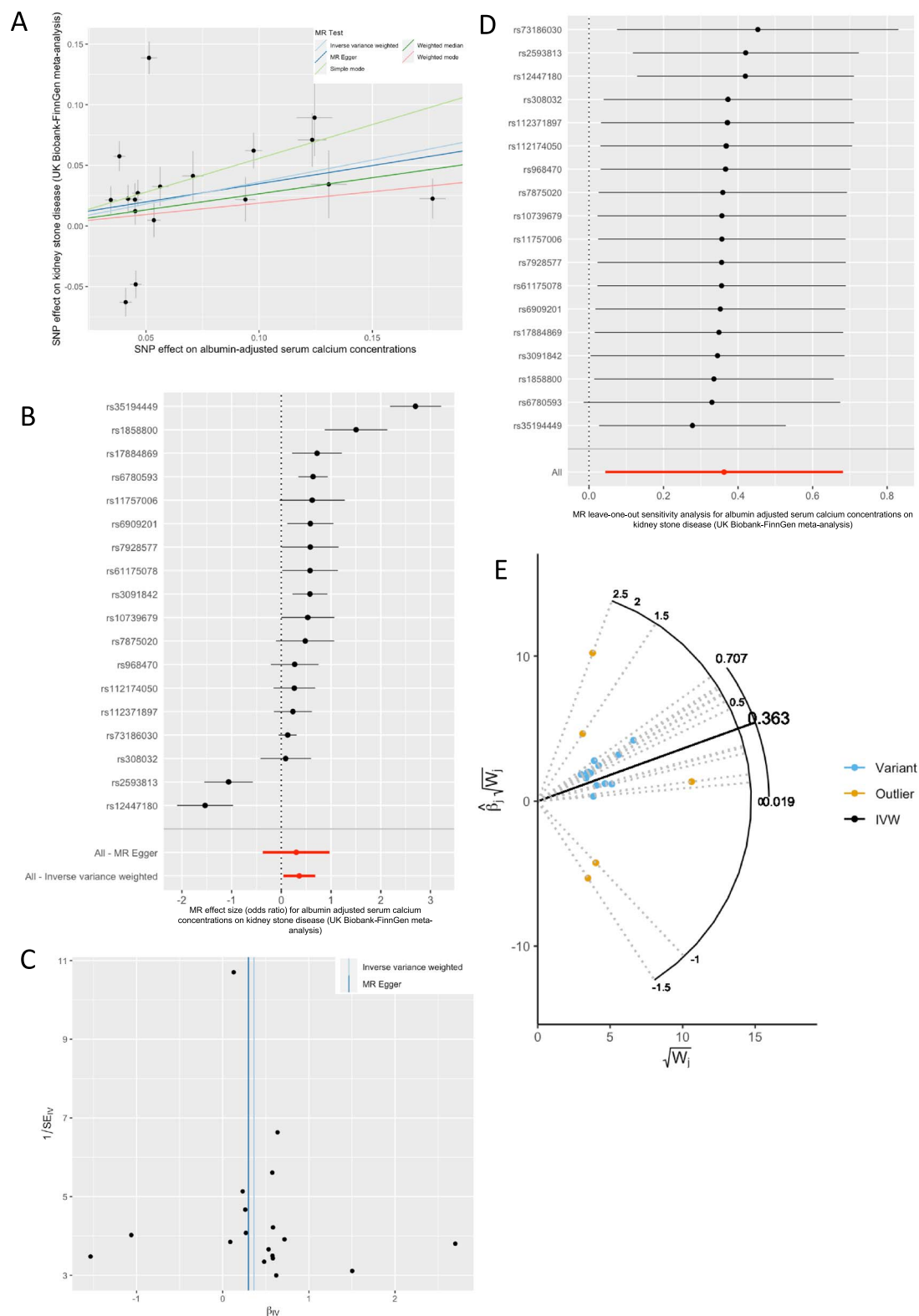


Figure 5. Visual representations of MR analyses of the effect of albumin-adjusted serum calcium concentrations on risk of kidney stone disease—reanalysis of data from Lovegrove et al.¹⁹ MR, Mendelian randomization; SNP, single nucleotide polymorphism. (A) Scatter plot of genetic associations and causal estimates. (B) Forest plot of variant-specific causal estimates, IVW, and MR-Egger estimates. (C) Funnel plot of variant-specific causal estimates. The estimate effect is plotted against the precision (reciprocal of standard error) of the estimate for each genetic variant. The vertical lines correspond to the estimates for IVW and MR-Egger estimates. (D) Forest plot of leave one out analyses. For each variant, the effect of removing it from the IV is plotted. (E) Radial plot. The x-axis represents the weight attributed to each genetic variant and the y-axis represents its Z-statistic (point estimate divided by standard error). The greater the vertical distance between a genetic variant and the overall causal estimate, the greater its contribution to Cochran's Q statistic of heterogeneity.

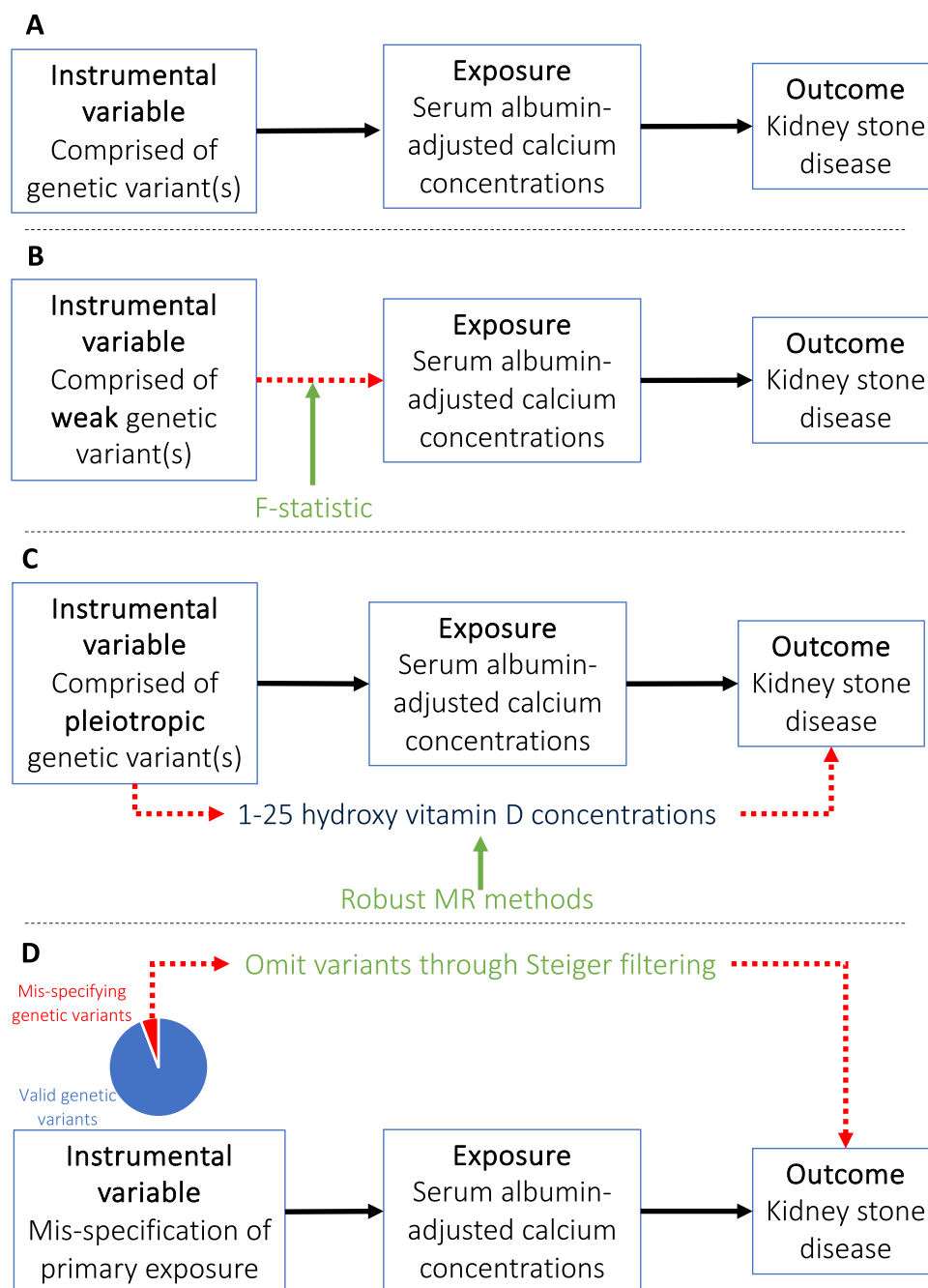


Figure 6. Assessing assumptions in MR. (A) Assumption in MR: MR assumes that there is an IV comprised of genetic variant(s) which are a proxy of the effect of an exposure variable in an outcome measure. (B) The F-statistic can evaluate the strength of association between the IV and exposure variable. (C) Robust MR methods, for example MR-Egger regression, can identify when genetic variants have pleiotropic effects and produce a reliable effect estimate so long as the magnitude of pleiotropic effects is not correlated with the variant-exposure effects. (D) Steiger filtering selects genetic variants that explain more variance in the exposure variable than the outcome and excludes those that fail to meet this condition to mitigate against mis-specification of the primary exposure.

of BMI.⁹⁶ Findings suggested that effects were bidirectional, and that increasing heel BMD increased liability to knee and hip osteoarthritis. Multivariable MR indicated that there is a BMI-independent pathway by which higher BMD causally increases liability to osteoarthritis.⁹⁶

Serum calcium and BMD

Cerani *et al.* aimed to determine whether higher serum calcium concentrations are linked to higher BMD and protection against osteoporotic fractures.⁹⁷ Using summary statistics

from a serum calcium concentrations GWAS and a GWAS meta-analysis of fractures including data from GEFOS they found no evidence for an effect of serum calcium on BMD or fracture risk.^{43,98} The authors acknowledge that the study may not reflect the effect of calcium supplementation on these outcomes. These results align with those of Trajanoska *et al.* who evaluated the role of 15 risk factors, including estimated calcium intake, on liability to osteoporotic fracture using MR and concluded that BMD was the only one of the evaluated risk factors to show a causal effect.⁴³ Moreover,

findings from MR studies corroborate with those from the RECORD (Randomized Evaluation of Calcium Or vitamin D) RCT which did not show that routine oral supplementation with calcium reduces fracture risk.⁹⁹

SGLT2-inhibition and BMD

Finally, Dai *et al.* used drug-target MR to ascertain whether there may be a beneficial effect from sodium-glucose cotransporter 2 (SGLT2) inhibition on BMD, risk of osteoporosis, or fracture.¹⁰⁰ SGLT2 inhibitors have been employed as anti-diabetic agents and have proven to be beneficial in patients with cardiac or renal failure regardless of diabetic status; however, concern exists regarding increased fracture risk.^{101–105} The authors constructed an IV for SGLT2 inhibition based on genetic variants that affect *SLC5A2* expression, have regional-wide association with HbA1c ($p < 1 \times 10^{-4}$), and explain genetic colocalization between *SLC5A2* expression and HbA1c concentrations.¹⁰⁰ Using this IV, Dai *et al.* found no evidence for a causal effect of SGLT2 inhibition on fracture risk and recommend that these results can be used to inform guidelines for patients prescribed this class of medications. Their results support those of RCTs reporting no effect of SGLT2 inhibitors on markers of bone formation, resorption, BMD, or fracture risk.^{106–110}

Software in MR

Multiple MR software packages are at researchers' disposal and can be implemented in standard statistics packages, for example R and Stata. Software packages typically include methods to derive several MR estimates, perform sensitivity analyses, and assumption testing. Software packages for MR also enable researchers to derive graphs as a visual representation of MR results. The MendelianRandomization package, implemented in R, can extract summarized data related to associations with named exposure and outcome variables from PhenoScanner (<http://www.phenoscaner.medschl.cam.ac.uk/>), a curated database of publicly-available results from large-scale GWAS, hosted by the University of Cambridge.^{111,113} Similarly, the TwoSampleMR package, also implemented in R, enables researchers to derive exposure or outcome IVs from the University of Bristol's Integrative Epidemiology Unit Open GWAS Project (<https://gwas.mrcieu.ac.uk/>) comprising a manually-curated collection of complete GWAS summary datasets, available as open source files, that can be queried or downloaded.^{31,49,90} The same group have also published MR-Base (<https://app.mrbase.org/>), a web app aiming to simplify MR implementation.^{29,45} As well as searching for genetic variants across multiple GWAS, it can perform 2-sample MR, sensitivity analyses, and produce MR visualization plots.

Reporting guidelines in MR studies

In 2021, the "STrengthening the Reporting of OBservational studies in Epidemiology- Mendelian randomization" (STROBE-MR) guidelines were published, providing MR studies with reporting guidelines similar to those already in existence for cohort, case-control, and cross-sectional studies.^{1,15,114} Akin to the original STROBE checklist, STROBE-MR recommends using a 20-point checklist (available at <https://www.strobe-mr.org/>) to enable readers,

journal reviewers, and editors to evaluate the quality of an MR study. The checklist is divided into sections of a paper (Title and abstract; Introduction; Methods; Results; Discussion; Other information) with opportunity for authors to detail the page location in their manuscript and the specific text that addresses the relevant checklist item.

Future directions: pragmatic use of MR studies

MR offers many practical advantages over traditional observational and interventional studies, for example deriving results rapidly rather than waiting for months or years for outcomes to arise in conventional epidemiological research, and the facility to draw on large, publicly-available GWAS datasets to inform clinical trials, thus improving the cost-efficiency of primary research in healthcare. However, caution is needed to ensure that the findings from MR studies are interpreted appropriately (Box 1).

Box 1. Considerations when appraising MR literature.

Study design:

- Is there a logical rationale for the study?
- Might the study contribute to the scientific literature?

Instrumental variables:

- Are the data from which IVs have been derived relevant, well-powered, and of a suitable genetic ancestry?
- What p -value threshold has been used for inclusion of genetic variants in IVs?
- Has linkage-disequilibrium (LD) been accounted for (eg, by clumping variants or incorporating an LD matrix)?
- How strong is the IV (F-statistic, R^2)?

Analyses:

- Have a range of sensitivity analyses been considered including those that address pleiotropy?
- Is there consistency in results across sensitivity analyses?
- Have authors corrected for multiple testing?
- Have reasonable efforts been made to demonstrate that the 3 core assumptions (*relevance, independence, exclusion restriction*) of MR have not been violated?

Interpretation:

- Have results been triangulated with other data (eg, eQTL or pQTL data) and other study designs (conventional observational, interventional, in vitro, in vivo, *in silico*)?
- Do the conclusions align with the data that have been presented?

Beyond appraising the methods, discussed in this review, one should triangulate MR analyses results, evaluating and contextualizing the place that they have amidst evidence from other sources. Care should be taken to ensure that the MR study is testing a biologically-plausible hypothesis; IVs are valid genetic proxies; included data are derived from phenotypes, interventions, and study populations reflecting

the underlying research question (Box 1)⁶; MR analyses are reported in a reproducible manner; and that findings from other study designs support the arising MR results, or vice versa.¹¹⁵ Just as researchers critically appraise observational and interventional studies, the same rigor should be applied to the MR literature; STROBE-MR guidelines can be beneficial in this regard.^{14,15}

Conclusions

MR studies can complement conventional epidemiological methods to enable researchers to infer causal effects in health and disease. Novel approaches to MR analyses are actively evolving and can provide insights into disease pathophysiology by overcoming many of the limitations that surround observational and interventional methods. However, despite the benefits that MR can confer, MR is subject to limitations and evaluating the assumptions made is crucial when performing and interpreting MR studies.

Author contributions

Catherine E. Lovegrove (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing—original draft, Writing—review & editing), Sarah A. Howles (Conceptualization, Writing—review & editing), Dominic Furniss (Conceptualization, Writing—review & editing), and Michael V. Holmes (Conceptualization, Writing—review & editing)

Supplementary material

Supplementary material is available at *Journal of Bone and Mineral Research* online.

Funding

Work was supported by grants from Kidney Research UK (RP_030_20180306) to S.A.H. and D.F., The Urology Foundation to S.A.H., National Institute for Health Research (NIHR) Oxford Biomedical Research Centre to D.F., and the Wellcome Trust to S.A.H (204826/z/16/z). C.E.L. is an M.R.C. Clinical Research Training Fellow. S.A.H. is a Wellcome Trust Clinical Career Development Fellow.

Conflicts of interest

M.V.H. is an employee of 23andMe and holds stock in 23andMe.

Data availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Rights retention statement

For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission.

References

- Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Int J Surg*. 2014;12(12):1500-1524. <https://doi.org/10.1016/j.ijsu.2014.07.014>
- Fewell Z, Davey Smith G, Sterne JAC. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2007;166(6):646-655. <https://doi.org/10.1093/aje/kwm165>
- Smith GD, Ebrahim S. Data dredging, bias, or confounding: they can all get you into the BMJ and the Friday papers. *BMJ*. 2002;325(7378):1437-1438. <https://doi.org/10.1136/bmj.325.7378.1437>
- Zabor EC, Kaizer AM, Hobbs BP. Randomized controlled trials. *Chest*. 2020;158(1):S79-S87. <https://doi.org/10.1016/j.chest.2020.03.013>
- Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research. *BJOG Int J Obstet Gynaecol*. 2018;125(13):1716. <https://doi.org/10.1111/1471-0528.15199>
- Sobczyk MK, Zheng J, Davey Smith G, Gaunt TR. Systematic comparison of Mendelian randomisation studies and randomised controlled trials using electronic databases. *BMJ Open*. 2023;13(9):e072087. <https://doi.org/10.1136/bmjopen-2023-072087>
- Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a trialist? Deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiology*. 2017;28(5):653-659. <https://doi.org/10.1097/EDE.0000000000000699>
- Davey Smith G, Holmes MV, Davies NM, Ebrahim S. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *Eur J Epidemiol*. 2020;35(2):99-111. <https://doi.org/10.1007/s10654-020-00622-7>
- Mendel G. Experiments in plant hybridization (1865). *Verhandlungen Naturforschenden Ver Brünn*. Published online 1996, Accessed January 1, 2013 Ww Mendelweb OrgMendel Html.
- Miko I. Gregor Mendel and the principles of inheritance. *Nat Educ*. 2008;1(1):134
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89-R98. <https://doi.org/10.1093/hmg/ddu328>
- OpenMR Project. Background. Accessed June 17, 2024. <https://open-mr-project.github.io/background.html>
- MR Publications over Time. Accessed June 17, 2024. <https://explocomputer.github.io/PubmedPlot/articles/mr.html>
- Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *JAMA*. 2021;326(16):1614-1621. <https://doi.org/10.1001/jama.2021.18236>
- 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. <https://doi.org/10.1136/bmj.n2233>
- Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. *Nat Rev Methods Primer*. 2022;2:6. <https://doi.org/10.1038/s43586-021-00092-5>
- Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008;9(6):477-485. <https://doi.org/10.1038/nrg2361>
- Davey SG. Epigenesis for epidemiologists: does evo-devo have implications for population health research and practice? *Int J Epidemiol*. 2012;41(1):236-247. <https://doi.org/10.1093/ije/dys016>
- Lovegrove CE, Bešević J, Wiberg A, et al. Central adiposity increases risk of kidney stone disease through effects on serum calcium concentrations. *J Am Soc Nephrol*. 2023;34(12):1991-2011. <https://doi.org/10.1681/ASN.0000000000000238>
- Fenton TR, Lyon AW, Eliasziw M, Tough SC, Hanley DA. Phosphate decreases urine calcium and increases calcium balance: a

- meta-analysis of the osteoporosis acid-ash diet hypothesis. *Nutr J*. 2009;8(1):41. <https://doi.org/10.1186/1475-2891-8-41>
21. Holmes MV, Smith GD. Revealing the effect of CETP inhibition in cardiovascular disease. *Nat Rev Cardiol*. 2017;14(11):635-636. <https://doi.org/10.1038/nrcardio.2017.156>
 22. Labrecque JA, Swanson SA. Interpretation and potential biases of Mendelian randomization estimates with time-varying exposures. *Am J Epidemiol*. 2019;188(1):231-238. <https://doi.org/10.1093/aje/kwy204>
 23. Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey SG. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: Mendelian randomisation study. *BMJ*. 2020;369:m1203. <https://doi.org/10.1136/bmj.m1203>
 24. Zhao Q, Wang J, Spiller W, Bowden J, Small DS. Two-sample instrumental variable analyses using heterogeneous samples. *Stat Sci*. 2019;34(2):317-333. <https://doi.org/10.1214/18-STS692>
 25. Chen L, Smith GD, Harbord RM, Lewis SJ. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLoS Med*. 2008;5(3):e52. <https://doi.org/10.1371/journal.pmed.0050052>
 26. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res*. 2023;4:186. <https://doi.org/10.12688/wellcomeopenres.15555.3>
 27. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584-591. <https://doi.org/10.1038/s41588-019-0379-x>
 28. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006;17(4):360-372. <https://doi.org/10.1097/01.ede.0000222409.00878.37>
 29. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol*. 2013;42(4):1134-1144. <https://doi.org/10.1093/ije/dyt093>
 30. Buxton J, Turney J. *The Rough Guide to Genes & Cloning*. Penguin; 2007.
 31. Elsworth B, Lyon M, Alexander T, et al. The MRC IEU OpenGWAS data infrastructure. *Genetics*. 2020. <https://doi.org/10.1101/2020.08.10.244293>
 32. CARDIoGRAMplusC4D. CARDIoGRAMplusC4D Consortium. Accessed November 28, 2023. <http://www.cardiogramplusc4d.org/>
 33. DIAGRAM Consortium. Accessed November 28, 2023. <https://diagram-consortium.org/index.html>
 34. EGG (Early Growth Genetics) Consortium. Accessed November 28, 2023. <http://egg-consortium.org/>
 35. Home. Accessed November 28, 2023. <http://www.gefos.org/?q=home>
 36. GIANT Consortium. Giant Consortium. Accessed November 28, 2023. https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium
 37. GLGC. Accessed November 28, 2023. <http://www.lipidgenetics.org/>
 38. dbGaP Study. Accessed November 28, 2023. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000585.v2.p1
 39. Tastad C. IIBDGC. IIBDGC. Accessed November 28, 2023; <https://www.ibdgenetics.org/>
 40. Magic Investigators - Home Page. Accessed November 28, 2023. <https://magicinvestigators.org/>
 41. PGC - Psychiatric Genomics Consortium. Accessed November 28, 2023. <https://pgc.unc.edu/>
 42. Home. SSGAC. Accessed November 28, 2023. <https://www.the.ssgac.org>
 43. Trajanoska K, Morris JA, Oei L, et al. Assessment of the genetic and clinical determinants of fracture risk: genome wide association and Mendelian randomisation study. *BMJ*. 2018;362:k3225. <https://doi.org/10.1136/bmj.k3225>
 44. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol*. 2011;40(3):755-764. <https://doi.org/10.1093/ije/dyr036>
 45. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol*. 2016;40(7):597-608. <https://doi.org/10.1002/gepi.21998>
 46. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey SG. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies1. *Am J Clin Nutr*. 2016;103(4):965-978. <https://doi.org/10.3945/ajcn.115.118216>
 47. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res*. 2017;26(5):2333-2355. <https://doi.org/10.1177/0962280215597579>
 48. Hemani G, Bowden J, Davey SG. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet*. 2018;27(R2):R195-R208. <https://doi.org/10.1093/hmg/ddy163>
 49. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenotype. *elife*. 2018;7:e34408. <https://doi.org/10.7554/eLife.34408>
 50. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol*. 2016;45(6):dyw220-dyw1974. <https://doi.org/10.1093/ije/dyw220>
 51. 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(2):512-525. <https://doi.org/10.1093/ije/dyv080>
 52. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol*. 2017;32(5):377-389. <https://doi.org/10.1007/s10654-017-0255-x>
 53. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304-314. <https://doi.org/10.1002/gepi.21965>
 54. 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(6):1985-1998. <https://doi.org/10.1093/ije/dyx102>
 55. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693-698. <https://doi.org/10.1038/s41588-018-0099-7>
 56. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol*. 2015;181(4):251-260. <https://doi.org/10.1093/aje/kwu283>
 57. 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(3):713-727. <https://doi.org/10.1093/ije/dyy262>
 58. Carter AR, Sanderson E, Hammerton G, et al. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *Eur J Epidemiol*. 2021;36(5):465-478. <https://doi.org/10.1007/s10654-021-00757-1>
 59. Gill D, Zuber V, Dawson J, et al. Risk factors mediating the effect of body mass index and waist-to-hip ratio on cardiovascular outcomes: Mendelian randomization analysis. *Int J Obes*. 2021;45(7):1428-1438. <https://doi.org/10.1038/s41366-021-00807-4>
 60. Burgess S, Daniel RM, Butterworth AS, Thompson SG, the EPIC-InterAct Consortium. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation

- in causal pathways. *Int J Epidemiol*. 2015;44(2):484-495. <https://doi.org/10.1093/ije/dyu176>
61. Burgess S. Violation of the constant genetic effect assumption can result in biased estimates for non-linear Mendelian randomization. *Hum Hered* Published online. 2023;88(1):79-90. <https://doi.org/10.1159/000531659>
 62. Staley JR, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genet Epidemiol*. 2017;41(4):341-352. <https://doi.org/10.1002/gepi.22041>
 63. Sulc J, Sjaarda J, Kutalik Z. Polynomial Mendelian randomization reveals non-linear causal effects for obesity-related traits. *Hum Genet Genomics Adv*. 2022;3(3):100124. <https://doi.org/10.1016/j.xhgg.2022.100124>
 64. Sofianopoulou E, Kaptoge SK, Afzal S, et al. RETRACTED: estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses. *Lancet Diabetes Endocrinol*. 2021;9(12):837-846. [https://doi.org/10.1016/S2213-8587\(21\)00263-1](https://doi.org/10.1016/S2213-8587(21)00263-1)
 65. Butler-Laporte G, Richards JB. Mendelian randomisation and vitamin D: the importance of model assumptions. *Lancet Diabetes Endocrinol*. 2023;11(1):14-15. [https://doi.org/10.1016/S2213-8587\(22\)00342-4](https://doi.org/10.1016/S2213-8587(22)00342-4)
 66. Wade KH, Carslake D, Sattar N, Davey Smith G, Timpson NJ. BMI and mortality in UK biobank: revised estimates using Mendelian randomization. *Obesity*. 2018;26(11):1796-1806. <https://doi.org/10.1002/oby.22313>
 67. Wade KH, Hamilton FW, Carslake D, Sattar N, Davey Smith G, Timpson NJ. Challenges in undertaking nonlinear Mendelian randomization. *Obesity*. 2023;31(12):2887-2890. <https://doi.org/10.1002/oby.23927>
 68. Daghlas I, Gill D. Mendelian randomization as a tool to inform drug development using human genetics. *Camb Prisms Precis Med*. 2023;1:e16. <https://doi.org/10.1017/pcm.2023.5>
 69. Schmidt AF, Finan C, Gordillo-Marañón M, et al. Genetic drug target validation using Mendelian randomisation. *Nat Commun*. 2020;11(1):3255. <https://doi.org/10.1038/s41467-020-16969-0>
 70. Stein EA, Mellis S, Yancopoulos GD, et al. Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. *N Engl J Med*. 2012;366(12):1108-1118. <https://doi.org/10.1056/NEJMoa1105803>
 71. Navarese EP, Kołodziejczak M, Schulze V, et al. Effects of proprotein convertase subtilisin/kexin type 9 antibodies in adults with hypercholesterolemia: a systematic review and meta-analysis. *Ann Intern Med*. 2015;163(1):40-51. <https://doi.org/10.7326/M14-2957>
 72. Besseling J, Kastelein JJ, Defesche JC, Hutten BA, Hovingh GK. Association between familial hypercholesterolemia and prevalence of type 2 diabetes mellitus. *JAMA*. 2015;313(10):1029-1036. <https://doi.org/10.1001/jama.2015.1206>
 73. White J, Sverdlow DI, Preiss D, et al. Association of lipid fractions with risks for coronary artery disease and diabetes. *JAMA Cardiol*. 2016;1(6):692-699. <https://doi.org/10.1001/jama.cardio.2016.1884>
 74. Fall T, Xie W, Poon W, et al. Using genetic variants to assess the relationship between circulating lipids and type 2 diabetes. *Diabetes*. 2015;64(7):2676-2684. <https://doi.org/10.2337/db14-1710>
 75. Schmidt AF, Sverdlow DI, Holmes MV, et al. PCSK9 genetic variants and risk of type 2 diabetes: a Mendelian randomisation study. *Lancet Diabetes Endocrinol*. 2017;5(2):97-105. [https://doi.org/10.1016/S2213-8587\(16\)30396-5](https://doi.org/10.1016/S2213-8587(16)30396-5)
 76. de Carvalho LSF, Campos AM, Sposito AC. Proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors and incident type 2 diabetes: a systematic review and meta-analysis with over 96,000 patient-years. *Diabetes Care*. 2018;41(2):364-367. <https://doi.org/10.2337/dc17-1464>
 77. Sabatine MS, Leiter LA, Wiviott SD, et al. Cardiovascular safety and efficacy of the PCSK9 inhibitor evolocumab in patients with and without diabetes and the effect of evolocumab on glycaemia and risk of new-onset diabetes: a prespecified analysis of the FOURIER randomised controlled trial. *Lancet Diabetes Endocrinol*. 2017;5(12):941-950. [https://doi.org/10.1016/S2213-8587\(17\)30313-3](https://doi.org/10.1016/S2213-8587(17)30313-3)
 78. Cao YX, Liu HH, Dong QT, Li S, Li JJ. Effect of proprotein convertase subtilisin/kexin type 9 (PCSK9) monoclonal antibodies on new-onset diabetes mellitus and glucose metabolism: a systematic review and meta-analysis. *Diabetes Obes Metab*. 2018;20(6):1391-1398. <https://doi.org/10.1111/dom.13235>
 79. Ference BA. How to use Mendelian randomization to anticipate the results of randomized trials. *Eur Heart J*. 2018;39(5):360-362. <https://doi.org/10.1093/eurheartj/ehx462>
 80. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26(9):1205-1210. <https://doi.org/10.1093/bioinformatics/btq126>
 81. Pendergrass SA, Brown-Gentry K, Dudek SM, et al. The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genet Epidemiol*. 2011;35(5):410-422. <https://doi.org/10.1002/gepi.20589>
 82. Schmidt AF, Holmes MV, Preiss D, et al. Phenome-wide association analysis of LDL-cholesterol lowering genetic variants in PCSK9. *BMC Cardiovasc Disord*. 2019;19(1):240. <https://doi.org/10.1186/s12872-019-1187-z>
 83. Gill D, Georgakis MK, Walker VM, et al. Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res*. 2021;6:16. <https://doi.org/10.12688/wellcomeopenres.16544.1>
 84. Wang Q, Oliver-Williams C, Raitakari OT, et al. Metabolic profiling of angiotensin-like protein 3 and 4 inhibition: a drug-target Mendelian randomization analysis. *Eur Heart J*. 2021;42(12):1160-1169. <https://doi.org/10.1093/eurheartj/ehaa972>
 85. Minicocci I, Montali A, Robciuc MR, et al. Mutations in the ANGPTL3 gene and Familial combined Hypolipidemia: a clinical and biochemical characterization. *J Clin Endocrinol Metab*. 2012;97(7):E1266-E1275. <https://doi.org/10.1210/jc.2012-1298>
 86. Ruscica M, Zimetti F, Adorni MP, Sirtori CR, Lupo MG, Ferri N. Pharmacological aspects of ANGPTL3 and ANGPTL4 inhibitors: new therapeutic approaches for the treatment of atherogenic dyslipidemia. *Pharmacol Res*. 2020;153:104653. <https://doi.org/10.1016/j.phrs.2020.104653>
 87. Sylvers-Davie KL, Davies BSJ. Regulation of lipoprotein metabolism by ANGPTL3, ANGPTL4, and ANGPTL8. *Am J Physiol - Endocrinol Metab*. 2021;321(4):E493-E508. <https://doi.org/10.1152/ajpendo.00195.2021>
 88. Spitler KM, Shetty SK, Cushing EM, Sylvers-Davie KL, Davies BSJ. Chronic high-fat feeding and prolonged fasting in liver-specific ANGPTL4 knockout mice. *Am J Physiol-Endocrinol Metab*. 2021;321(4):E464-E478. <https://doi.org/10.1152/ajpendo.00144.2021>
 89. Bowden J, Spiller W, Del Greco MF, 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(4):1264-1278. <https://doi.org/10.1093/ije/dyy101>
 90. Hemani G, Tilling K, Smith GD. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet*. 2017;13(11):e1007081. <https://doi.org/10.1371/journal.pgen.1007081>
 91. Cho Y, Haycock PC, Sanderson E, et al. Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework. *Nat Commun*. 2020;11(1):1010. <https://doi.org/10.1038/s41467-020-14452-4>

92. Turin TC, Okamura T, Afzal AR, et al. Hypertension and lifetime risk of stroke. *J Hypertens*. 2016;34(1):116-122. <https://doi.org/10.1097/HJH.0000000000000753>
93. Hopewell JC, Clarke R. Emerging risk factors for stroke: what have we learned from Mendelian randomization studies? *Stroke*. 2016;47(6):1673-1678. <https://doi.org/10.1161/STROKEAHA.115.010646>
94. Sanderson E, Richardson TG, Hemani G, Davey SG. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *Int J Epidemiol*. 2021;50(4):1350-1361. <https://doi.org/10.1093/ije/dyaa288>
95. Hartley A, Hardcastle SA, Paternoster L, et al. Individuals with high bone mass have increased progression of radiographic and clinical features of knee osteoarthritis. *Osteoarthritis Cartil*. 2020;28(9):1180-1190. <https://doi.org/10.1016/j.joca.2020.03.020>
96. Hartley A, Sanderson E, Granell R, et al. Using multivariable Mendelian randomization to estimate the causal effect of bone mineral density on osteoarthritis risk, independently of body mass index. *Int J Epidemiol*. 2022;51(4):1254-1267. <https://doi.org/10.1093/ije/dyab251>
97. Cerani A, Zhou S, Forgetta V, et al. Genetic predisposition to increased serum calcium, bone mineral density, and fracture risk in individuals with normal calcium levels: Mendelian randomization study. *BMJ*. 2019;366:l4410. <https://doi.org/10.1136/bmj.l4410>
98. Morris JA, Kemp JP, Youlten SE, et al. An atlas of genetic influences on osteoporosis in humans and mice. *Nat Genet*. 2019;51(2):258-266. <https://doi.org/10.1038/s41588-018-0302-x>
99. Oral vitamin D3 and calcium for secondary prevention of low-trauma fractures in elderly people (randomised evaluation of calcium or vitamin D, RECORD): a randomised placebo-controlled trial. *Lancet*. 2005;365(9471):1621-1628. [https://doi.org/10.1016/S0140-6736\(05\)63013-9](https://doi.org/10.1016/S0140-6736(05)63013-9)
100. Dai H, Zheng L, Zhu Z, et al. Evaluation of the effect of sodium-glucose cotransporter 2 inhibition on fracture risk: evidence from Mendelian randomization and genetic association study. *J Bone Miner Res*. 2023;38(11):1645-1653. <https://doi.org/10.1002/jbmr.4880>
101. Empagliflozin in patients with chronic kidney disease. *N Engl J Med*. 2023;388(2):117-127. <https://doi.org/10.1056/NEJMoa2204233>
102. Heerspink HJL, Stefánsson BV, Correa-Rotter R, et al. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med*. 2020;383(15):1436-1446. <https://doi.org/10.1056/NEJMoa2024816>
103. Solomon SD, McMurray JJV, Claggett B, et al. Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction. *N Engl J Med*. 2022;387(12):1089-1098. <https://doi.org/10.1056/NEJMoa2206286>
104. McMurray JJV, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med*. 2019;381(21):1995-2008. <https://doi.org/10.1056/NEJMoa1911303>
105. Packer M, Anker SD, Butler J, et al. Cardiovascular and renal outcomes with Empagliflozin in heart failure. *N Engl J Med*. 2020;383(15):1413-1424. <https://doi.org/10.1056/NEJMoa2022190>
106. Bolinder J, Ljunggren Ö, Johansson L, et al. Dapagliflozin maintains glycaemic control while reducing weight and body fat mass over 2 years in patients with type 2 diabetes mellitus inadequately controlled on metformin. *Diabetes Obes Metab*. 2014;16(2):159-169. <https://doi.org/10.1111/dom.12189>
107. Wiviott SD, Itamar R, Bonaca MP, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2019;380(4):347-357. <https://doi.org/10.1056/NEJMoa1812389>
108. Ljunggren Ö, Bolinder J, Johansson L, et al. Dapagliflozin has no effect on markers of bone formation and resorption or bone mineral density in patients with inadequately controlled type 2 diabetes mellitus on metformin. *Diabetes Obes Metab*. 2012;14(11):990-999. <https://doi.org/10.1111/j.1463-1326.2012.01630.x>
109. Ruanpeng D, Ungprasert P, Sangtian J, Harindhanavudhi T. Sodium-glucose cotransporter 2 (SGLT2) inhibitors and fracture risk in patients with type 2 diabetes mellitus: a meta-analysis. *Diabetes Metab Res Rev*. 2017;33(6):e2903. <https://doi.org/10.1002/dmrr.2903>
110. Li X, Li T, Cheng Y, et al. Effects of SGLT2 inhibitors on fractures and bone mineral density in type 2 diabetes: an updated meta-analysis. *Diabetes Metab Res Rev*. 2019;35(7):e3170. <https://doi.org/10.1002/dmrr.3170>
111. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics*. 2016;32(20):3207-3209. <https://doi.org/10.1093/bioinformatics/btw373>
112. Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics*. 2019;35(22):4851-4853. <https://doi.org/10.1093/bioinformatics/btz469>
113. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol*. 2017;46(6):1734-1739. <https://doi.org/10.1093/ije/dyx034>
114. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335(7624):806-808. <https://doi.org/10.1136/bmj.39335.541782.AD>
115. Porcu E, Gilardi F, Darrous L, et al. Triangulating evidence from longitudinal and Mendelian randomization studies of metabolomic biomarkers for type 2 diabetes. *Sci Rep*. 2021;11(1):6197. <https://doi.org/10.1038/s41598-021-85684-7>