PRIMER

Mendelian randomization

Eleanor Sanderson[1,2✉], M. Maria Glymour[3], Michael V. Holmes[1,4,5], Hyunseung Kang[6], Jean Morrison[7], Marcus R. Munafò[1,8,9], Tom Palmer[1,2], C. Mary Schooling[10,11], Chris Wallace[12,13], Qingyuan Zhao[14] and George Davey Smith[1,2,9]

Abstract | Mendelian randomization (MR) is a term that applies to the use of genetic variation to address causal questions about how modifiable exposures influence different outcomes. The principles of MR are based on Mendel's laws of inheritance and instrumental variable estimation methods, which enable the inference of causal effects in the presence of unobserved confounding. In this Primer, we outline the principles of MR, the instrumental variable conditions underlying MR estimation and some of the methods used for estimation. We go on to discuss how the assumptions underlying an MR study can be assessed and describe methods of estimation that are robust to certain violations of these assumptions. We give examples of a range of studies in which MR has been applied, the limitations of current methods of analysis and the outlook for MR in the future. The differences between the assumptions required for MR analysis and other forms of epidemiological studies means that MR can be used as part of a triangulation across multiple sources of evidence for causal inference.

Instrumental variable
(IV). A variable associated with an exposure that is not associated with the outcome through any other pathway.

Natural experiment
Natural experiments are variation in any exposures or risk factors that occurred by chance in the population without conscious or deliberate intervention from investigators or scientists.

Mendelian randomization (MR) uses genetic variation to address causal questions about whether modifiable exposures influence health, developmental or social outcomes[1]. Exposures can be any factor robustly associated with genetic variation in individuals; for example, an exposure could include measurable characteristics of an individual such as body mass index (BMI) or less directly observable traits such as the expression of a particular gene in a specific tissue.

The statistical methodology for MR is generally based on instrumental variable (IV) analysis. An IV, or 'instrument', is related to the exposure but not to the outcome of interest, other than through its association with the exposure. IV analysis was first proposed a century ago and is an approach to causal inference that uses an IV to make causal effect estimates in the presence of unobserved confounding of the exposure and the outcome. IV analyses can be applied to any source of variation in an exposure that is unrelated to the outcome, including investigator-initiated treatment randomization in a randomized controlled trial (RCT) or when a natural experiment provides a plausible source of exogenous or unconfounded variation[2–4]. MR is based on the assumption that genetic variants provide a source of such exogenous variation in the exposure and can therefore act as an IV[1]. MR can be applied using any genetic variation that satisfies the requirements of an IV[5], although it is usually implemented using single-nucleotide polymorphisms (SNPs). BOX 1 further outlines the principles of MR.

Using genetic variants in this way, MR avoids bias from unobserved confounding of the exposure and outcome. However, there are important additional assumptions required for causal inference and effect estimation that are different to those used in other causal effect estimation methods. Causal effect estimates from MR can be evaluated within a triangulation of evidence framework, which involves interpreting findings alongside results from complementary approaches that rely on different assumptions. When using this approach, it is important that sources of bias in different study modalities are unrelated to each other so that the magnitude and direction of the bias in one study will not predict the size and direction of bias in the others[6–8].

MR studies — especially two-sample studies using previously published summary-level genetic association data — provide a rapid and affordable approach to evaluating causal questions. There is an urgent need for these tools because many causal questions in health research cannot be adequately answered with conventional observational study designs and are not amenable to evaluation with RCTs for logistical or ethical reasons. MR is especially appealing because it relies on assumptions that differ from those of conventional observational studies and therefore circumvents some of their common biases[1,8]. The range of applications of MR and closely related methods for understanding causal mechanisms has increased rapidly in the past 20 years. The increasing availability of data and the vast expansion of IV methods have overcome some of the original barriers to MR caused by lack of data and the inability to assess the robustness of results obtained[1]. Major investments in collecting genetic data within large research studies have

✉e-mail: *eleanor.sanderson@bristol.ac.uk*

https://doi.org/10.1038/s43586-021-00092-5

Author addresses

[1]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
[2]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.
[3]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA.
[4]MRC Population Health Research Unit, University of Oxford, Oxford, UK.
[5]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK.
[6]Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA.
[7]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA.
[8]School of Psychological Science, University of Bristol, Bristol, UK.
[9]National Institute for Health Research (NIHR), Biomedical Research Centre, University of Bristol, Bristol, UK.
[10]School of Public Health, Li Ka Shing, Faculty of Medicine, The University of Hong Kong, Hong Kong, China.
[11]School of Public Health, City University of New York, New York, NY, USA.
[12]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
[13]Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), University of Cambridge, Cambridge, UK.
[14]Statistical Laboratory, University of Cambridge, Cambridge, UK.

enabled numerous applications of MR and allowed for increased statistical power and more precise effect estimates. Further, methodological innovation to enhance MR analyses is flourishing and innovations aim to allow for correct estimation with more plausible assumptions and estimate more complex effects, which include independent effects of multiple phenotypes or age-sensitive exposures. We therefore focus on the principles of MR and detail a few core MR estimation methods. The methods for MR listed here should not be taken as a definitive list of all potential methods available.

In this Primer, we provide guidance on the underlying principles of MR, discuss the information necessary to decide whether an MR approach is appropriate and feasible, and review best contemporary practices for MR. We outline the principles and assumptions underlying MR, along with the data required. Next, we detail the core methods for estimation of causal effects and explain how the assumptions underlying MR can be verified or subjected to sensitivity analyses. We then describe a range of studies that have applied MR in different settings, detail the importance of triangulating MR results with findings using other study designs and discuss steps to improving the openness of research involving MR. Finally, we outline sources of bias that may affect MR studies that cannot be corrected for with current methods and discuss some of the challenges and opportunities for MR in the future.

Experimentation

The essence of an MR design is that the association between a genetic variant (G) and an outcome (Y) can be used to test whether and by how much the exposure of interest (X) influences the outcome, provided that the genetic variant is associated with the exposure of interest and has no other source of association with the outcome[1,8,9]. Bias originating from confounding of the exposure and outcome should not influence the MR estimate. The rationale of MR studies parallels that of RCTs in which randomization influences the treatment received by participants, there are no confounders of randomization and the outcome and has no other plausible mechanism to influence health outcomes other than through treatment (FIG. 1). In RCTs, randomly assigned treatment therefore evaluates the effect of treatment on the outcome, whereas in MR, a genetic variant is treated as a naturally occurring form of randomization.

Confounder
A trait that influences both the exposure and outcome of interest.

As an example, FIG. 2a shows a directed acyclic graph for an RCT aimed at estimating the causal effect of lowering circulating levels of the inflammatory marker C-reactive protein (CRP) on systolic blood pressure (SBP), in which participants are randomized to receive a CRP-lowering medication or placebo. Alternatively, the effect of long-term differences in circulating CRP could be estimated with MR by considering a genetic variant that is known to alter CRP levels (FIG. 2b). The directed acyclic graphs for both studies are the same as long as certain assumptions are satisfied (discussed below).

In our hypothetical RCT, an intention-to-treat analysis can be conducted to determine whether the treatment influences the outcome by comparing SBP among individuals randomly assigned to the CRP-lowering medication to SBP in participants randomly assigned to placebo[10,11]. Intention-to-treat analysis estimates the effect on the outcome of being assigned to the group allocated to treatment, rather than receiving that treatment. A frequently used approach for analysis is to compare the mean SBP among individuals randomized to treatment to the mean SBP among individuals randomized to control:

$$\beta_1 = E(\text{SBP}|G=1) - E(\text{SBP}|G=0) \quad (1)$$

where β_1 is the effect on SBP of being assigned to the treatment group, G is an indicator of randomization and SBP is measured systolic blood pressure. Alternatively, a linear regression can be used:

$$E(\text{SBP}|G) = \beta_0 + \beta_1 G \quad (2)$$

where β_0 is a constant. As there are no confounders of randomization and SBP, there is no need to control for any variables to derive an unconfounded estimate of the effect of randomization. Therefore, in a setting where G is binary, β_1 as estimated in Eq. 1 is identical to β_1 as estimated in Eq. 2 and both estimate the causal effect of randomized treatment groups on SBP. Being randomized to CRP-lowering medication should only affect SBP if there is a causal effect of CRP on SBP.

A potential disadvantage of the intention-to-treat estimate, for many questions of substantive interest, is that it does not give the magnitude of the effect of the exposure on the outcome — for example, of CRP on SBP in the above example. It only determines whether or not there is a causal effect. To estimate the size of that causal effect, the degree to which the instrument affects the exposure must be taken into account. IV analyses are an alternative estimation method that can be used to derive an estimate of the causal effect of the treatment (here, CRP) on the outcome (SBP) by accounting for the size of the association between randomization and CRP[3,4,12–15]. In this scenario, randomization becomes the instrument for the estimation. In its simplest form, IV analysis takes the ratio of the effect of randomization on SBP to the effect of randomization on CRP:

$$\gamma_1 = \frac{E[\text{SBP}|G=1] - E[\text{SBP}|G=0]}{E[\text{CRP}|G=1] - E[\text{CRP}|G=0]} \qquad (3)$$

where γ_1 is known as the Wald ratio estimator and CRP is the level of circulating C-reactive protein. The numerator of Eq. 3 is simply Eq. 1, but here the association is scaled by the effect of randomization on CRP. Under the IV conditions described in BOX 2, this estimator provides a test of whether there is a causal effect of CRP on SBP.

IV analyses can be applied to any potential source of randomization, including intentionally designed RCTs or quasi-randomization in natural experiments[15,16]. The term MR is applied when the quasi-randomization arises from genetic variation and a phenotype influenced by the genetic variant is the exposure of interest[17,18]. The genetic variant is referred to as the genetic instrument. For example, naturally occurring genetic variants in the gene encoding CRP regulate blood levels of CRP and such variants have been used to estimate the effects of circulating CRP levels on SBP[19,20].

The above example highlights an important difference between RCTs and MR: RCTs estimate the effect of a particular intervention or treatment over the timeframe of the study, whereas MR estimates the lifetime effects of the genetic variants, as discussed in a recent preprint[21]. This can lead to substantial differences in the effect estimates obtained, owing to the differences in the time period over which the effects are estimated. There are a number of other differences between RCTs and MR. Although MR was first proposed using family data where the difference in alleles between siblings is random, data limitations mean that most MR is conducted using data on unrelated individuals[1]. In MR using unrelated individuals, the similarity between the allele groups is not guaranteed as it is in a well-conducted RCT. Further, associations between allele distribution and traits can exist at a population level owing to population stratification or assortative mating. The particular genetic variants used in the MR may also have effects on the outcome that are not due to the exposure received by the individual[22]. These issues all represent violations of the conditions required for IV estimation, which are described in detail below. How these violations may occur in MR studies and potential mechanisms to detect such violations are discussed in the 'Results' and 'Limitations and optimizations' sections of this Primer.

Box 1 | **The principles of the MR approach**

The Mendelian randomization (MR) approach draws on Mendel's first and second laws of genetic inheritance: the law of segregation and the law of independent assortment[206]. The law of segregation indicates that at every point in the autosomal genome, offspring randomly inherit one allele from their mother and one allele from their father. The law of independent assortment implies that these alleles will be passed to offspring independently of each other, except in regions of the genome that are genetically linked in the DNA of the parents.

The first extended exposition of MR[1] was in the context of family-based studies. Its analogy with randomized controlled trials was in the context of the random allocation of variants from parents to their children. At the time of this first description, adequate family-based data were not available and 'approximate' MR in population studies was advocated for instead; indeed, family-based data are still only used in a small minority of published MR studies. The advocacy of population studies was based on the premise that at a population level, genetic variants can identify groups that differ, on average, with respect to a modifiable exposure. In these studies, genetically defined group membership should be unrelated to factors that may confound conventional observational associations, including behavioural, social and physiological exposures that occur after conception[1,4,6,206,207]. Therefore, genetic associations between traits should be free from confounding and any difference in outcomes between groups defined by genetic variation can be attributed to the genetic variation, assuming no selection bias owing to that genetic variation.

Conditions required for MR estimation

Interpretation of results from MR studies relies on four conditions[12,23]. The first three of these conditions are usually referred to as the conditions for a valid IV and are required for any IV analysis to test whether the exposure has a causal effect on the outcome. These are described in BOX 2. In our simplified example of CRP and SBP, we imagine only a single IV; however, MR is easily extendable to take advantage of multiple genetic variants that influence the same exposure[24]. When multiple genetic variants can be identified that fulfill the IV conditions, they can be used to improve the statistical power of MR analyses[25,26].

The three IV conditions described in BOX 2 are sufficient to test the exact null hypothesis as they can determine the presence or lack of a causal effect of the exposure on the outcome. However, they are not sufficient to derive a point estimate of the size of the effect of the exposure on the outcome[27,28]. This requires an additional condition[27] known as a point-estimate-identifying condition or fourth IV condition. Several alternative point-estimate-identifying conditions — which permit subtly different interpretations of the IV estimate — have been described and researchers can adopt the version of the condition which seems most plausible for the setting at hand[17,28]. BOX 3 outlines the most popular of these alternative point-estimate-identifying conditions and the effect estimate obtained from each one. Additionally, the vast majority of MR estimation methods (with non-linear MR[29] being the notable exception) impose the assumption that the relationship between the exposure and the outcome is linear across different values of the exposures.

Biases that compromise the interpretation of an RCT can also undermine MR studies. For example, if random assignment in an RCT influences who participates in follow-up assessments, typical analyses of the RCT are biased. Similarly, if the genetic variants used in MR influence who has available outcome data — either owing to differential survival or study participation — the MR study will be biased[30].

Finally, data used in MR additionally require the assumption that changes in genetic variation are equivalent in their effects to changes in the exposure through environmental or pharmaceutical manipulation — a concept known as gene–environment equivalence[31]. Given that genetic variants will influence the developing human from conception, the interpretation is applied to the influence of the variants from conception onwards. These particular MR-related issues are discussed in BOX 4.

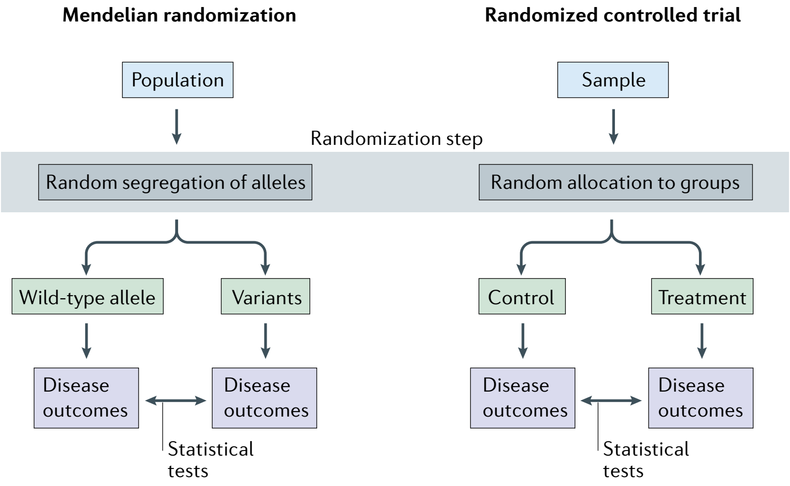


Fig. 1 | **An overview of MR studies.** This overview compares and contrasts the parallels between Mendelian randomization (MR) and randomized controlled trials (RCTs). In MR, randomization is due to the random allocation of alleles. This conceptualization was originally based on between-sibling variation, where allocation of alleles is random and not dependent on population-level variation (see also BOX 1). Inference from MR in this way relies on the assumption of gene–environment equivalence — that a change in the exposure caused by genetic variation has the same effect on the outcome as a change in that exposure caused by environmental factors.

Data used for MR estimation

MR studies can be conducted using individual-level data (including genetic and phenotype measures for each individual in the study) or summary data (on the association between each genetic instrument and the exposure and the outcome phenotypes of interest). Summary data are often obtained from genome-wide association studies (GWAS), which estimate the association between SNPs and the exposure and SNPs and the outcome traits.

When individual-level data are used for estimation, the statistical power of an MR analysis (or, equivalently, the precision of the estimate that can be derived) increases in proportion to the sample size and the variance in the exposure explained by the genetic instruments. When summary data are used, the precision of the MR estimate depends on how precisely the associations between the genetic variants and each of the exposure and the outcome have been estimated — in other words, how large the standard error of the estimated association is. Genetic variants typically explain only a small proportion of the variation in the relevant phenotype; as a result, low statistical power and imprecise effect estimates are common in MR studies and well-powered studies usually require large datasets. Power calculators are available for simple MR studies to determine whether a particular sample size is sufficient for the estimation to give reasonably precise results[32–35]. Simulation studies to determine power are also usually used to accommodate unique data features[36].

The association of the proposed genetic instrument with the exposure can be estimated in a sample other than that used to estimate the effect of the proposed genetic instrument on the outcome[37]. MR conducted in this way is referred to as 'two-sample MR'. The capacity to use two different samples for MR analyses has dramatically broadened the scope of MR studies because when either the desired exposure or outcome of a study is rare or expensive to measure, it can be difficult to identify a dataset with data on the genetic instrument, exposure and outcome. An important assumption for two-sample MR estimation is that the two samples are from the same underlying population, or more narrowly that the association between the genetic variants and exposure is the same in both samples, although that exposure may not be measured or reported in the sample included in the outcome dataset[38]. To satisfy this assumption, two-sample approaches usually use data from the most similar populations possible, with respect to genetic ancestry and contextual factors such as the prevalence of environmental exposures and the timeframe in which the measurements were taken.

The method of estimation and applicable sensitivity analyses used in MR depend on whether individual participant or summary-level data are used to conduct the analyses[39]. Using multiple genetic instruments in combination improves statistical power because the combination increases the total fraction of the exposure variance explained by the instruments[26,40]. The availability of multiple genetic instruments is also valuable for detecting or avoiding bias if one or more of the IV conditions are not met for some or all of the instruments.

Instrument selection

Genetic variants used as instruments for MR should be associated with the exposure of interest, so that they satisfy IV condition 1 (BOX 2). This can be through the use of variants with known functionality or through the selection of variants that are robustly associated with the exposure. GWAS can potentially identify a large number of SNPs that predict a selected phenotype and many MR studies use SNPs identified in credible GWAS as genome-wide significant predictors of the exposure of interest for estimation, that is those SNPs associated with the exposure with $P < 5.0 \times 10^{-8}$ (REF.[41]).

When using individual data, overlap between the dataset used for instrument discovery and the dataset

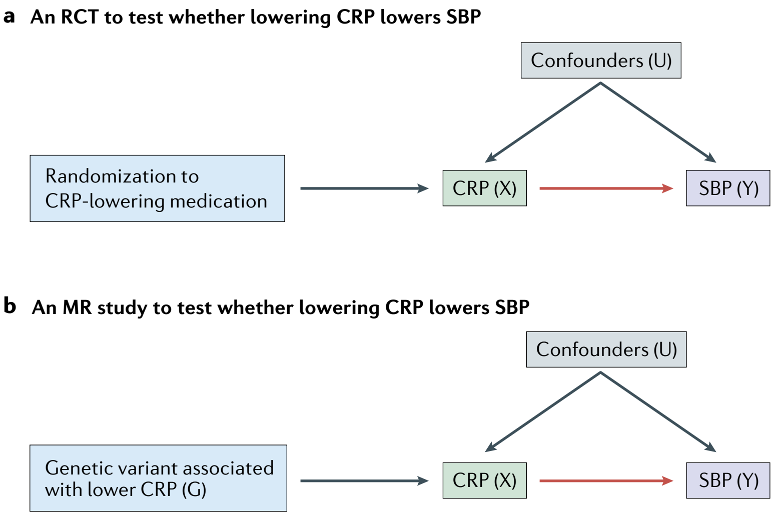


Fig. 2 | **Illustration of a randomized control study and instrumental variable estimation.** A randomized controlled trial (RCT) (panel **a**) and a Mendelian randomization (MR) study (panel **b**) to estimate the effect of lowering C-reactive protein (CRP) on systolic blood pressure (SBP). The arrows highlighted in red show the causal effect of interest.

Box 2 | **Instrumental variable conditions**

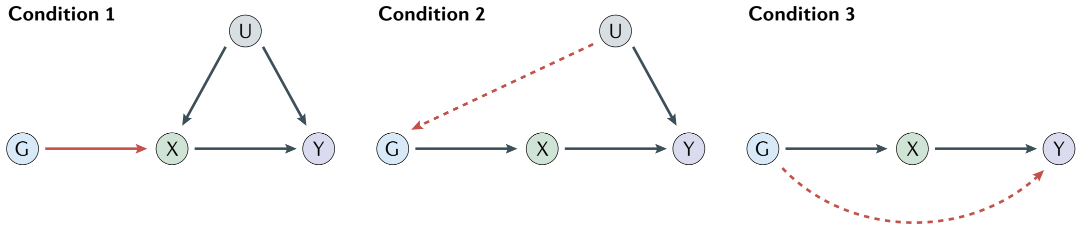
The instrumental variable (IV) conditions are required to hold for the results from any IV estimation — including a randomized controlled trial or Mendelian randomization (MR) estimation — to provide a valid test of the null hypothesis that the exposure has no effect on the outcome[12,17,23,52,208].

One way that the IV conditions can be expressed formally is with directed acyclic graphs (see the figure)[17]; solid red lines show effects that must exist and dashed red lines represent effects that must not exist if an IV is to be used to assess the causal effect of X on Y. G is the IV (a genetic variant or set of genetic variants in MR). U represents unobserved confounders. We do not consider here the potential bias owing to selection.

The IV conditions are as follows.

- IV condition 1: relevance. The IV is associated with the exposure.
- IV condition 2: exchangeability. There are no causes of the IV that also influence the outcome through mechanisms other than the exposure of interest (no confounders of the IV and the outcome).
- IV condition 3: the exclusion restriction. The IV does not affect the outcome other than through the exposure and does not affect any other trait that has a downstream effect on the outcome of interest.

Only the first condition can be formally tested. The other two conditions can be disproved and otherwise assessed through a range of sensitivity analyses, but cannot be demonstrated to be true[66,209]. Methods for testing the first condition and of assessing the plausibility of the second and third conditions are discussed in the 'Results' section.



used for estimation can introduce a bias known as 'winner's curse'. The goal of IV is to remove the effect on the exposure of variation due to confounders of the exposure and outcome. However, the best fitting model for the association of a SNP and the exposure will, by chance, pick up some variation owing to confounders. Although this bias is small and unimportant if the SNP has a very strong effect on the exposure, this is rarely the case. When many SNPs are used as IVs, each with a very small effect, this can create a non-trivial bias towards the conventional effect estimate, known as weak instrument bias[42]. This can be avoided through bias correction calculations or by using a two-sample approach and applying jackknife resampling to the estimation[43–45]. In a jackknife estimation, the data are divided into groups and each is then used for estimation, with instrument discovery conducted in the rest of the sample. The results for each group are then meta-analysed to obtain a result for the whole dataset (see preprint[46]).

Bias due to overfitting is a concern when summary-level data are used for estimation if the effect of the SNP on the exposure is in a dataset that overlaps with the dataset used to estimate the SNP–outcome association. Recent research has suggested that overlap between the samples used may not bias the results obtained by as much as previously thought, unless the instruments are not strongly associated with the exposure, and methods have been proposed to estimate the size of this bias and to correct for it[43] (see preprints[44,47]).

Results

This section outlines methods used for MR estimation, tests for violation of the first IV condition and methods of estimation that are robust to particular violations of the second and third IV conditions. Here, we cover the main methods used for estimation. A number of other papers are available that cover guidelines for reading[48], conducting[39] and interpreting[49] results from MR studies. STROBE guidelines for the consistent reporting of MR studies have also been published[50,51]. Additionally, the MR dictionary provides an extensive glossary of terms used in MR.

Individual-level data

Estimating causal effects. When using individual level data in MR estimation, genetic variants can either be used as separate instruments or combined into an allele score[25]. An allele score is generated by adding up the number of risk-increasing alleles for all the variants selected as instruments. This score can be unweighted, so that each SNP makes the same contribution, or weighted, so that the number of risk-increasing alleles at each SNP is multiplied by the estimated effect of that SNP on the exposure[25]. Weighted scores provide increased instrument strength and power, although there are cases in which the unweighted approach is preferable — for example, if the definition of the exposure in the discovery dataset differs from the exposure variable in the estimation data. In such a case the weights will reflect the weight of the SNP on an exposure that is not the exposure included in the estimation. The more similar the definition of the exposure is in each sample the more preferable the weighted approach will be. Differences in scaling alone will not affect the preference for a weighted score. Ideally, both SNPs and weights should be selected from a dataset that does not overlap with the dataset used to obtain the MR estimates, such as those from GWAS in non-overlapping datasets[52]. If many SNPs that each have only a small effect on the exposure are being used, combining them into a single score can

Box 3 | **Point-estimate-identifying conditions**

The instrumental variable (IV) conditions described in BOX 2 are sufficient to test for the presence of a causal effect. However, performing estimation and interpretation of the causal effect requires at least one additional assumption. The effect of the exposure (X) on the outcome (Y) may differ for different people. These differences require additional assumptions to be placed on the relationship between the instruments, exposure and outcome to identify both the causal effect of the exposure on the outcome, and to whom that causal effect estimate applies. Each assumption gives a slightly different interpretation for the causal effects obtained from Mendelian randomization (MR) analysis.

There are two frequently used asumptions for point-estimate-identifying conditions. The first option is homogeneity of the effect of the exposure on the outcome, or that either (a) the effect of the exposure on the outcome is the same for everyone, regardless of the starting value of X or any other individual characteristics, or (b) the effect of the exposure on the outcome does not depend on the value of the instrument. Option (a) gives the interpretation that the causal effect estimate is ‘the causal effect of the exposure on the outcome’, whereas option (b) gives the interpretation that the effect estimate obtained is the ‘population average of the causal effect of the exposure on the outcome’. The second assumption is monotonicity in the association between the genetic variants and the exposure — that the direction of the effect of the genetic variant on the exposure is the same for everyone[2,27,210–212]. This gives the interpretation that the effect estimate is the effect of the exposure on the outcome in those people whose exposure is changed by the instrument. In MR, this is the average effect of differences in the exposure that are attributable to differences in the genetic variants. For continuous exposures or outcomes, violation of the IV conditions allowing point identification can be accessed through examination of the variance of the trait by the level of the instrument (either per-allele or in a binary dominant model, as appropriate). Violations will lead to differences in the variance of the trait across the level of the instrument[213].

Which assumption is most relevant will depend on the particular estimation; however, the assumption of monotonicity is usually relevant for MR estimation. The point-estimate-identifying condition remains an area of debate and methodology development, with researchers identifying additional possible assumptions that would support a causal interpretation of the IV effect estimate.

---

increase the power of the analysis and reduce the risk of bias from many weak instruments[26]. However, if any SNPs violate IV conditions 2 or 3 (if any of the component SNPs influence the outcome through a mechanism other than the exposure of interest) then the allele score will also violate that condition.

Estimation of causal effects using individual-level data is usually implemented with some version of two-stage least-squares (2SLS) estimation (alternative methods include likelihood approaches that are common in structural equation modelling)[53]. 2SLS estimation for MR uses genetic variants to obtain a predicted value of the exposure ($\widehat{X}$) that is not associated with any of the unmeasured confounders. The first stage can be written as:

$$X = \pi_0 + \mathbf{G}\boldsymbol{\pi} + v_x \qquad (4)$$

where X is the exposure of interest; $\mathbf{G}$ is a $n \times L$ matrix of genetic variants, where n is the number of individuals in the dataset and L is the number of SNPs; $\boldsymbol{\pi}$ is a vector of the effect of each genetic variant on the exposure of length L; π_0 is a constant and V_x is a random error term. The outcome is then regressed upon the predicted value of the exposure, $\widehat{X}$:

$$Y = \alpha + \beta\widehat{X} + u \qquad (5)$$

where Y is the outcome, α is a constant, β is the effect of the exposure on the outcome and u is a random error term assumed to be unrelated to v_x. The four conditions for IV estimation imply that the assumption of independence of u and v_x is met and the estimated value of β — that is, $\widehat{\beta}$, obtained from Eq. 5 — is a consistent estimator for the effect of X on Y. If the estimation is implemented using an allele score, Eq. 4 is replaced with:

$$X = \pi_0 + \pi S + v_x \qquad (6)$$

where S is the allele score (weighted or unweighted) and π is a single coefficient for the association of the genetic score with the exposure. The second stage of the analysis, Eq. 5, is the same whether we are using individual SNPs as instruments or an allele score. In both cases, the standard error should not be computed using the standard formula for linear models and should be corrected for the additional uncertainty owing to the inclusion of $\widehat{X}$ in the estimation. IV estimation software packages implement this correction as standard.

Additional measured covariates can be incorporated into both stages of the estimation. The use of additional covariates should be considered carefully because covariates can be influenced by the exposure or the outcome. In either of these situations, controlling for such a covariate could bias the MR effect estimate[54–56].

Assessment of IV conditions. Regardless of the statistical method being used, it is important to assess the IV conditions. The first IV condition can be tested using a first-stage *F* statistic, which tests the association between the SNPs and the exposure. If the genetic instruments are not strongly associated with the exposure, then weak instrument bias can be introduced into the estimation[42]. The first-stage *F* statistic should be reported in all MR analyses. As a general rule, if the first-stage *F* statistic is greater than 10, the level of this bias is small[57,58]. A cut-off of $F > 10$ has been used as a conventional threshold for a strong instrument in some studies. We note that this should not be used as a rigid rule and an *F* statistic <10 does not indicate that this instrument should not be used, rather that weak instrument bias should be considered as an issue in analysis.

Although the second and third IV conditions cannot be proved to be true, they can sometimes be disproved. Assessment of these conditions therefore focuses on disproving them, and failure to disprove the conditions is interpreted as supporting the validity of the proposed IV. Genetic variants are fixed at conception, so it is not possible for conventional confounders such as age, sex or environmental risk factors to influence them. However, confounding of the genetic variants with the outcome in a sample can be induced by population stratification, dynastic effects and assortative mating[59], violating the second IV condition. This confounding is not easily corrected with current MR methods and is discussed in more detail in the ‘Limitations and optimizations’ section.

Violations of the third IV condition can be caused by pleiotropy, where genetic variants have effects on multiple phenotypes[60,61]. This can include misspecification of the primary phenotype where the phenotype of interest is not the phenotype that the SNP is primarily

First-stage *F* statistic
Test statistic used to test the strength of association between the instrument(s) and the exposure in an instrumental variable estimation.

Linkage disequilibrium
Correlation between genetic variants located closely together on the genome.

associated with[8,61,62]. Additionally, linkage disequilibrium means that the effects of neighbouring genetic variants can introduce additional associations between the variant of interest — and thus the exposure it relates to — and the outcome, creating a bias analogous to that caused by pleiotropy. Pleiotropy in the context of MR is described in FIG. 3. Many MR methods are available that are robust to different forms of pleiotropy and analyses using these different methods should be carried out in any MR study to determine how sensitive the results are to an assumption of no pleiotropy.

A final important source of bias in MR, and indeed all studies of observational data, is selection bias[63,64]. This selection could occur either from differential selection into the sample or selection on a competing risk for the outcome. Selection bias cannot be accounted for easily with existing MR methods and is discussed further in the 'Limitations and optimizations' section.

An approach for assessing the IV assumptions that is applicable when there are more IVs than exposures of interest is based on over-identification tests. These tests, such as the Sargan test[65], leverage the expectation that if all proposed IVs are valid, they should deliver identical IV effect estimates. If the IV effect estimates from multiple IVs differ to a greater extent than expected owing to sampling error, then at least one is not valid for the exposure–outcome effect of interest. If all IVs are biased in the same way, over-identification tests will not identify the bias; for example, over-identification tests can incorrectly suggest a lack of pleiotropy even when it is present if similar pleiotropic pathways are likely to affect many or all proposed IVs or if there is population stratification biasing the association between many SNPs and the outcome in the same way[24]. They also rely on the assumption that each IV estimates the same causal effect, which may not be true for complex traits where different genetic variants potentially act as genetic instruments for different aspects of the trait. The weaker the effect of an IV on an exposure, the more imprecise the IV effect estimate will be and therefore the more likely it becomes that an instrument will fail to reject an over-identification test.

Box 4 | **Issues interpreting MR results**

Gene–environment equivalence

Typically, Mendelian randomization (MR) considers exposures that are modifiable and so evidence of a causal effect of the exposure on the outcome can be used to infer that an exposure intervention will lead to a change in the outcome. However, making such an inference depends on the exposure of interest fulfilling the consistency criterion that however the intervention is applied to alter the exposure, the effect on the outcome is the same. This means that changes in an exposure by either a hypothetical change in genotype or by a change in the environment should produce the same downstream effect on an outcome[31,214–216]. For example, genotypic influences on circulating cholesterol level or a similar change in cholesterol level induced by dietary influences should lead to the same effect on coronary heart disease. Although many exposures can be closely proxied by genetic variation, for others — such as those that reflect aspects of social deprivation and income — it is unlikely that genetic variation will mimic environment changes exactly[217]. Gene–environment equivalence is a fundamental principle in MR and consideration should be given to how likely it is to hold when interpreting the results from any MR study.

Interpretation of results for time-varying exposures

Genetic variants are fixed throughout an individual's lifetime and MR estimates can therefore be interpreted as the 'lifetime effect' of the exposure on the outcome[1,9]. If the association between the genetic variants and the exposure is constant across the life course, this lifetime effect can be interpreted as the effect of having a level of exposure that is a unit higher at every time point across the life course[218]. However, for many exposures the association between genetic variants and the exposure may vary across the life course; for example, genetic variants associated with body mass index have been shown to have a wide range of differential effects between childhood and adulthood[141]. In this scenario, MR estimates can be interpreted as the lifetime effect of being on a trajectory for the exposure associated with having an exposure level that is a unit higher at the time it is measured[21]. Multivariable MR can be used to estimate causal effects of the different time periods and potentially to identify particularly relevant periods across the life course[141,219]. That MR estimates the lifetime effect of the exposure on the outcome means that MR estimates can be larger than estimates obtained from alternative methods of estimation, such as randomized controlled trials, because the total length of time over which the exposure can have an effect is much longer.

One further method for identifying potential violations of the IV conditions when the exposure is binary or categorical is using IV inequality constraints[28,66,67]. The IV conditions described above imply a set of mathematical patterns that must be true if the conditions are true; these patterns can be used to demonstrate that the IV conditions are not met if the equalities defined by those patterns do not hold. IV inequalities are rarely especially informative because they identify only extreme violations of the conditions. These inequalities can also be used to define non-parametric bounds for an IV estimate (those that would hold without the fourth, point-estimate-identifying condition discussed above). Although these bounds are often very wide, they can give a sense of how much an IV analysis depends on the point-estimate-identifying condition. An alternative approach for identifying violations of the IV conditions is to examine the association between the genetic variants and other measured causes of the outcome, excluding any variables that are themselves on the same pathway as the exposure of interest (see preprint[68])[69]. If a proposed genetic instrument predicts other causes of the outcome that are not thought to be along the same causal pathway as the exposure, it indicates that the proposed instrument is not valid.

Methods, such as sisVIVE[70] and adaptive LASSO[71], provide MR estimates that are robust to pleiotropy under certain assumptions. These methods assume that multiple IVs are available and that a majority or plurality of the proposed IVs are valid. Given this assumption, it is possible to estimate the magnitude of pleiotropic bias. An alternative approach is to adjust for pleiotropic effects of the genetic variants by accounting for the association between the genetic variants and potentially pleiotropic phenotypes. Methods that apply this approach include constrained IVs[72] and multivariable MR[73].

Tests to invalidate proposed IVs often draw on subject matter knowledge, such as an understanding of settings in which a genetic variant does not influence the exposure, where the genetic variant may have different effects based on the level of an environmental variable (known as gene–environment interactions) or where the exposure should have no effect on the outcome, such as a negative control or zero-relevance point. The proposed genetic instrument should not be associated with the outcome in an environmental setting where it is not associated with the exposure unless there are pleiotropic pathways from the genetic variant to the outcome.

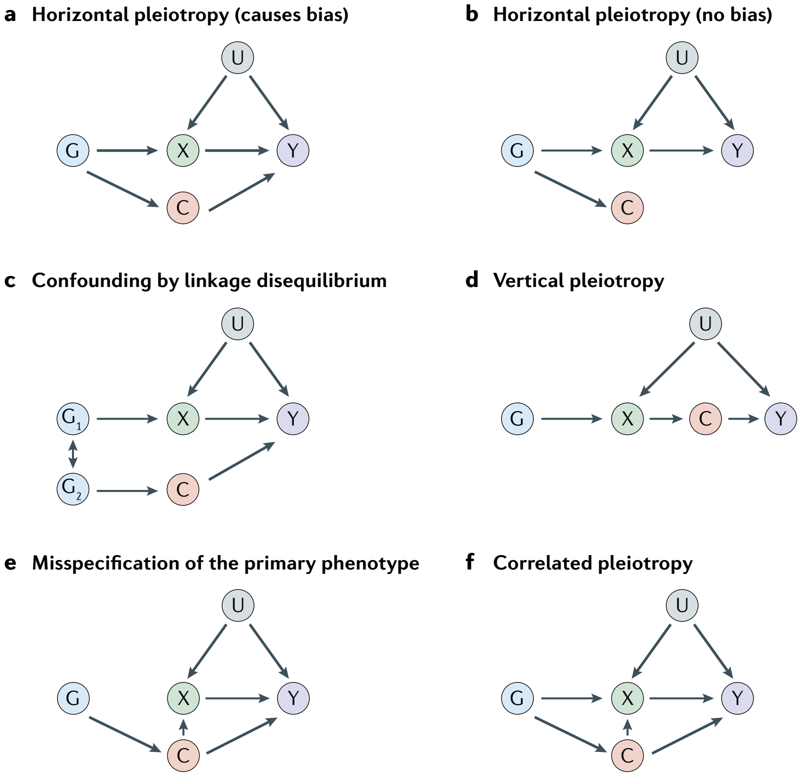


Fig. 3 | **Types of pleiotropy.** Figure showing different types of pleiotropy in Mendelian randomization (MR), where G is a genetic variant or set of genetic variants associated with the exposure, X is the exposure of interest, Y is the outcome of interest, U is an unmeasured confounder and C is another (potentially unmeasured) phenotype that is also associated with the genetic variants. **a**,**b** | Horizontal pleiotropy. Sometimes referred to as biological pleiotropy, this occurs where a genetic variant is associated with multiple phenotypes and these phenotypes lie on different pathways. In horizontal pleiotropy with bias (panel **a**), the third instrumental variable condition (IV3) is violated because there is a pathway from the genetic variant to the outcome that does not occur via the exposure. In horizontal pleiotropy with no bias (panel **b**), as the genetic variants are not associated with other phenotypes on the pathway to the outcome, MR estimates are not biased. **c** | Confounding by linkage disequilibrium. When G_2 has an effect on the outcome through a pathway that is not via the exposure, correlation between G_1 and G_2 creates a bias that is indistinguishable from that shown in panel **a**. **d** | Vertical pleiotropy. Another phenotype lies on the genetic variant–exposure–outcome pathway. This could occur either before or after the exposure of interest. Sometimes referred to as mediated pleiotropy, this form of pleiotropy does not bias MR studies and can even be used to elucidate causal intermediaries[41]. **e** | Misspecification of the primary phenotype. Vertical pleiotropy can bias MR estimates if the wrong phenotype is specified as the primary phenotype. Here the genetic variants are primarily associated with C. If X is misspecified as the primary phenotype, MR estimation of the effect of X on Y would be biased by the alternative pathways from C to Y[8,61]. **f** | In correlated pleiotropy, genetic variants for the exposure are also associated with a confounder of the exposure and outcome. In this setting, the size of the pleiotropic effect is correlated with the size of the association between the genetic variant and the exposure. This form of pleiotropy is particularly hard to detect and correct for. The scenarios in panels **b** and **d** produce settings where the pleiotropy will not bias the MR estimation. All other settings violate assumptions IV2 or IV3 and can cause meaningful bias in MR estimation.

Vertical pleiotropy
The phenomenon of a genetic variant associated with multiple phenotypes on the same pathway.

A classic example of this type of analysis is examining the effect of alcohol consumption in populations where subgroups of the population (for example, women in some cultures) do not drink or drink very little[74]. If the IV conditions are satisfied, there should be no association between genetic variants for alcohol consumption and the outcome under consideration among women in the previous example. Two methods, MR GxE and MR GENIUS, have extended and formalized these concepts and enable the estimation of causal effects in more general settings. MR GxE uses an interaction between the genetic variant and a covariate to create a new IV (see preprint[75])[76]; MR GENIUS uses variation that occurs owing to unobserved interactions between the genetic variants and covariates as the instrument[75,77].

Summary-level data

Estimating causal effects. MR estimation with summary level data requires estimates of $\widehat{\pi}_l$, the estimated effect of genetic variant l on the exposure with variance $\sigma^2_{x,l}$, and $\widehat{\Gamma}_l$, the estimated effect of genetic variant l on the outcome with variance $\sigma^2_{y,l}$. Inverse-variance weighting (IVW) estimation is a meta-analysis of the variant specific Wald ratios for each variant, which are given as:

$$\widehat{\beta}_l = \frac{\widehat{\Gamma}_l}{\widehat{\pi}_l}$$

where $\widehat{\beta}_l$ is the effect estimated using genetic variant l. These individual ratios are weighted by their associated uncertainty; the IVW estimator $\widehat{\beta}_{\mathrm{IVW}}$ can therefore be computed as:

$$\widehat{\beta}_{\mathrm{IVW}} = \frac{\sum_{l=1}^{L} \widehat{\pi}_l \widehat{\Gamma}_l \sigma_{y,l}^{-2}}{\sum_{l=1}^{L} \widehat{\pi}_l^2 \sigma_{y,l}^{-2}}$$

where L is the total number of genetic variants included as potential IVs[37]. The IVW estimate can equivalently be obtained by regressing the genetic variant–outcome association, $\widehat{\Gamma}_l$, on the genetic variant–exposure association, $\widehat{\pi}_l$, (without an intercept) weighted by the inverse variance of the SNP–outcome association ($1/\widehat{\sigma}^2_{y,l}$):

$$\widehat{\Gamma}_l = \beta_{\mathrm{IVW}} \widehat{\pi}_l + u_l \text{ weighted by } 1/\widehat{\sigma}^2_{y,l}$$

This equation describes a linear regression with the intercept fixed to zero as $u_l \sim N(0, 1)$, and is based on a dataset with L observations.

One important assumption for IVW estimation is that the genetic variants are independent of each other[40]. This assumption is usually satisfied by removing one of each pair of genetic variants that are in linkage disequilibrium. However, methods are available that can take linkage disequilibrium into account between genetic variants in summary-level MR[78,79]. It is also important to ensure that data are harmonized to ensure that the values of $\widehat{\Gamma}_l$ and $\widehat{\pi}_l$ refer to the same effect alleles[80].

Assessment of IV conditions. As with individual level data analysis, IV conditions need to be assessed for any summary-data MR. A number of different methods are available to correct for horizontal pleiotropy — a violation of the third IV condition — under different assumptions about the causal structure of that pleiotropy. TABLE 1 lists some of these methods, which primarily draw on three approaches: outlier removal, outlier adjustment and adjustment for specific forms of pleiotropy. Many methods combine more than one of

Horizontal pleiotropy
The phenomenon of a genetic variant associated with multiple phenotypes on different pathways.

these approaches. Outlier removal estimation involves the identification and removal of individual genetic variants for which the causal effect estimate obtained using that variant alone lies outside the expected range given the estimates obtained from other variants, so that they do not have an effect on the result obtained. Traditionally, summary-data MR is visualized as a scatter plot plotting associations of the variant and exposure against associations of the variant and outcome (FIG. 4a,b); however, this can limit the identification of outliers. Radial MR is a method for visualizing the data that can make outlying data points easier to detect[81] (FIG. 4c). An additional approach is to explore the effect of individual SNPs on the overall IV estimate, by methods such as leave-one-out analyses (FIG. 4d). Methods of estimation that use outlier removal include weighted median[82], weighted mode[83] and MR LASSO[84]. Outlier adjustment methods identify outlying variants and then perform an adjustment to either the effect obtained from that genetic variant or to the weight given to the estimate from that variant so that the variant has less influence on the overall estimation result. Many pleiotropy-robust MR methods fall into this category, including MR Tryx[85], MR PRESSO[86], MR Robust[84], MR RAPS[87], MR GRAPPLE[88] and MR CAUSE[89]. The final broad category of pleiotropy-robust methods for summary-data MR estimation are methods that allow for most or all of the genetic variants included in the estimation to have pleiotropic effects on the outcome and to place other constraints on the pleiotropic effects. These methods include MR Egger[90] and multivariable MR[73,91]. Each of these methods imposes strong assumptions on the nature of the pleiotropy. MR Egger analysis assumes that across all instruments, the magnitude of the pleiotropic effect is unrelated to the strength of the association between the genetic variant and the phenotype of interest (known as the InSIDE assumption). This assumption will not hold when there is correlated pleiotropy (FIG. 3f). Multivariable MR assumes that pleiotropic pathways operate through known phenotypes that are also included in the estimation.

None of the methods described above is truly robust to all types of pleiotropy and each imposes different assumptions on the nature of the pleiotropy and how the pleiotropic effects are accounted for. Furthermore, many methods have less statistical power than conventional MR, leading to very wide confidence intervals. Therefore, a few methods should be selected on the basis of the most plausible assumptions for the application in question and used alongside an IVW MR estimation to perform a sensitivity analysis; this can determine how robust MR results are to the assumption that genetic variants have no pleiotropic effects on the outcome under different alternative specifications. As a minimum, any summary-data MR estimation usually includes weighted median and weighted

Table 1 | **List of MR estimation methods**

Category	Core IV assumption relaxed	Individual-level data	Summary data
‘Basic’ MR method	None	Wald ratio estimation, 2SLS regression analysis[a]	Wald ratio estimation, IVW[a,37]
Weak instrument robust methods	IV1; allows for weak instruments	LIML[26], allele score approaches[26]	MR RAPS[87], debiased IVW[187], MR GRAPPLE[88], NOME adjustment[188], two-sample AR[189]
Outlier/variant selection and removal	IV3; allows for balanced/sparse pleiotropy	Weighted median[190]	Weighted median[a,82]
Outlier/variant selection and removal	IV3; allows for (some) directional pleiotropy	sisVIVE[70], adaptive LASSO[71], weighted mode[190]	Weighted mode[a,83], MR LASSO[84], Steiger filtering[a,93], Welch-weighted Egger[94], contamination mixture[191], GSMR[79], MR-Clust[192], Bayesian MIMR[193], CIV[72]
Outlier/variant adjustment	IV3; allows for balanced pleiotropy	Limited approaches currently available	MR RAPS[87], MRCIP[194]
Outlier/variant adjustment	IV3; allows for (some) directional pleiotropy	Limited approaches currently available	MR TRYX[85], MR Robust[84], MR CAUSE[89], MR PRESSO[86], MR GRAPPLE[88], MRMix[195], MR-LDP[196], IMRP[197], regularization[198], MR-PATH (see preprint[199])
Estimation adjustment	IV3; allows for balanced pleiotropy	Limited approaches currently available	Debiased IVW[187]
Estimation adjustment	IV3; allows for (some) directional pleiotropy	Constrained IVs[72], multivariable MR[73]	MR Egger[90], multivariable MR[73,91], MR Link[200], hJAM[201], GIV[202], Bayesian network analysis[203], BMRE[204], BayesMR[205]
Environmental control adjustment	IV3; allows for (some) directional pleiotropy	MR GxE[75,76], MR GENIUS[77]	Limited approaches currently available

2SLS, two-stage least-squares; IV, instrumental variable; LIML, limited information maximum likelihood; MR, Mendelian randomization. [a]Most frequently used methods; we note that each method relies on strong assumptions and may not be the most appropriate in any particular setting. These categories are not mutually exclusive and the classification of some methods may be ambiguous. Each method will impose some alternative version of the IV condition that is relaxed for consistent estimation with that method. Methods that are robust to directional pleiotropy impose (often strong) assumptions on the nature of that pleiotropy to enable estimation. Novel MR estimation methods are being developed continually and will generally fit into one or more of these categories.

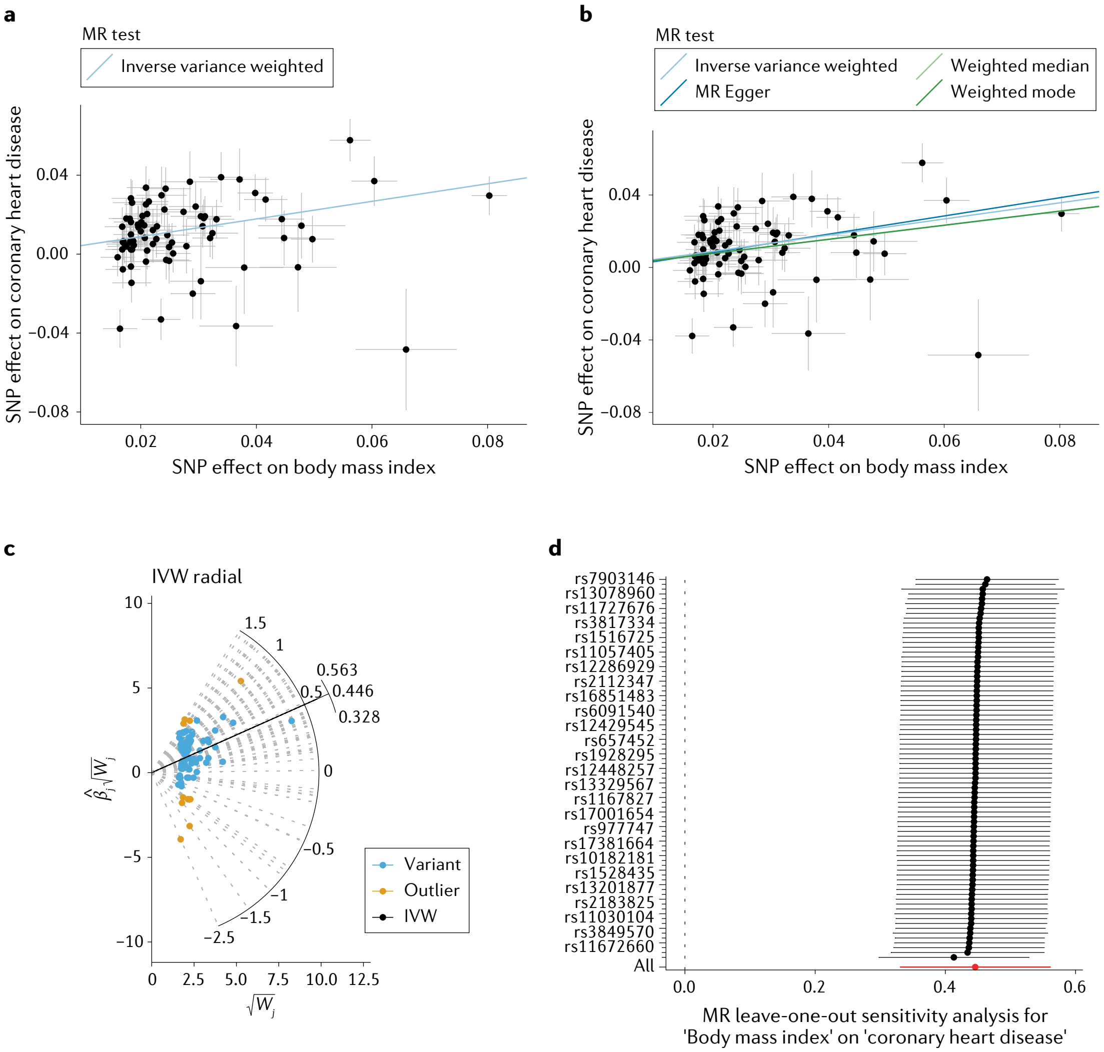


Fig. 4 | **Data visualization.** Figure showing different visualizations of a summary-data Mendelian randomization (MR) analysis. The example shown is estimating the effect of body mass index on coronary heart disease (CHD). **a** | A scatter plot of the single-nucleotide polymorphism (SNP)–exposure and SNP–outcome associations for each SNP with an inverse-variance weighting (IVW)-estimated line fitted. The error bars around each point show the standard error of the estimated association between the SNP and the exposure and the SNP and the outcome. **b** | The same plot with the robust approaches of weighted mode, weighted median and MR Egger added (note that the weighted median line is obscured by the weighted mode line). **c** | The same data plotted using a radial MR framework to identify outliers; the horizontal axis shows the weight given to each point and the vertical axis shows the weight multiplied by the effect estimate. The IVW-estimated fitted line is shown. **d** | A leave-one-out analysis where the IVW estimate has been recalculated, excluding one SNP at a time, to look for SNPs that highly influence the overall result. These graphs were created using the 'TwoSampleMR' and 'RadialMR' R packages, using data from the OpenGWAS project. Code used to create these figures is detailed in the Supplementary information for illustrative purposes.

mode approaches, although these can be replaced with appropriate alternatives for the application in question. Additionally, these estimation methods will not necessarily identify violations of any IV conditions that are not due to pleiotropy of the nature interrogated by the method. Consequently, consistent results across a range of methods is not a guarantee that results are free from bias. Potential violations of the IV assumptions not due to pleiotropy are discussed in the 'Limitations and optimizations' section.

Another form of pleiotropy arises when the exposure for the MR estimation is misspecified and genetic variants associated with a confounder are used as instruments for the exposure under investigation (FIG. 3e). For example, BMI influences circulating CRP and if a genetic variant primarily associated with BMI is included as a genetic variant for CRP, misleading effect estimates of the causal effect of CRP on other phenotypes — including BMI — can be generated[61,92]. These issues are increasingly important to consider because the sample sizes used in GWAS are increasing, making it more likely that a primary phenotype has been misspecified (in the context of GWAS, this could refer to the detection of genetic variants for an upstream phenotype of the exposure which potentially confounds the exposure and outcome, or genetic variants for the outcome if the direction of effect has been misspecified). Steiger filtering attempts to correct for this misspecification by removing SNPs that explain more variation in the outcome than the exposure[93]. Any genetic variant should explain more variation in the phenotypes that it is more proximal to; however, differing measurement error, substantially different sample sizes for each phenotype, or the presence of binary or categorical phenotypes can lead to phenotypes that are less proximal to the genetic variant appearing to have more variation explained by the variant than more proximal phenotypes in the observed data. Additional methods are now being developed that attempt to resolve misspecification and confounding[89,94,95].

Software packages

Any statistical package can be used for simple MR estimates as the core IV estimate is derived from a two-step regression model. Deriving correct standard errors requires special calculations and variations on the standard model have been implemented as packages in common statistics packages such as Stata and R. A range of software packages are available in both Stata and R to conduct MR estimation, many of which include a range of assumption tests and options to conduct robust methods. The TwosampleMR R package links to the OpenGWAS project database (see preprint[96]), a large database of GWAS results that can be used in the estimation. TABLE 2 gives details of the most popular software packages currently available; an extended list is given in the Supplementary information.

Further extensions of MR methods

Bidirectional MR. In bidirectional MR, two MR analyses are conducted on the same pair of phenotypes by reversing the exposure and the outcome. This method can be used to establish the direction of effect between two variables. For example, extensive observational evidence indicates that hearing loss predicts dementia and it is hypothesized to be an important causal determinant of dementia[97]; however, it is possible that the neurodegenerative disease that leads to dementia also causes hearing loss and thus the causal direction between hearing loss and dementia is unclear. There are known genotypes for both hearing loss and Alzheimer disease — the most common cause of dementia[98–100] — and a bidirectional MR would first conduct an MR analysis of the effect of liability to dementia on hearing and then for the effect of hearing on dementia. If genetic variants known to associate with dementia influence hearing loss and genetic variants known to associate with hearing loss do not influence dementia risk, this suggests that hearing loss is a causal determinant of dementia.

Results from bidirectional MR studies should be interpreted with caution. Evidence of an effect in both directions could indicate a true bidirectional relationship between the exposures or be a product of bias from horizontal pleiotropic effects in the variants, misspecification of the primary phenotype, or a violation of the second IV condition owing to confounding of genetic variants and outcome caused by factors such as population stratification and dynastic effects.

Bidirectional relationship
Where an effect acts in both directions between a pair of traits so that changing one will change the other.

Multivariable MR. Multivariable MR is an extension of standard MR that includes multiple exposures, predicted by a set of genetic variants used as instruments. FIGURE 5 illustrates a multivariable MR with two exposures. Although multiple exposures can be included in a multivariable MR, there must be at least as many genetic variants or scores included as instruments as there are exposures. Multivariable MR can be estimated with either individual-level or summary-level data using extensions of the 2SLS or IVW approaches, respectively[73,101]. Conditions required for estimation are adapted from the standard IV conditions and are defined as follows: each exposure must be robustly predicted by the instruments, conditional on the other exposures included in the estimation (multivariable instrumental variable condition 1, or MVIV1); there must be no confounders of the outcome and any of the instruments (MVIV2) and none of the instruments can have an effect on the outcome that does not act through at least one of the exposures (MVIV3). If the above conditions are met, the estimates obtained from multivariable MR will be a direct effect of each exposure included on the outcome, given the other exposures included in the estimation[73].

Multivariable MR can be used as an approach to address pleiotropic violations of the IV conditions. In a univariable MR where IV3 is violated and the genetic variants used as instruments for an exposure of interest are also thought to be associated with another trait on the path to the outcome, that trait can be included as an additional exposure in the multivariable MR estimation. Multiple, correlated exposures can be included in a multivariable MR; however, including multiple exposures can reduce power and potentially instrument strength and thus the benefit of adding extra exposures must be considered carefully. Bayesian approaches have been proposed for selecting a set of exposures where multiple highly correlated exposures are potentially relevant for an outcome[102]. In addition, multivariable MR can be used for mediation analysis, as described below.

MR mediation analysis. MR can be used to estimate the proportion of the effect of an exposure on an outcome that is mediated by an intermediate phenotype[103,104]. Network MR and two-step MR use two univariable MR estimates to do this, estimating the effect of the primary exposure on the intermediate phenotype and the effect of the intermediate phenotype on the outcome[105,106]. Alternatively, multivariable MR can estimate the direct effect of each exposure on the outcome that is not mediated by the other exposures included in the estimation. If all of the IV conditions are satisfied, this estimate will differ from a univariable MR estimate where all or part of the effect of the exposure on the outcome acts through a mediating phenotype included in the multivariable MR estimation[103]. Both two-step and multivariable MR can therefore be used as part of a mediation analysis to estimate how much of the effect of an exposure on an outcome acts through an intermediate phenotype[103,104]. When multiple intermediate phenotypes are thought to be potential mediators, two-step MR can estimate the proportion of the outcome mediated through each of these, whereas multivariable MR including all of the mediators considered will estimate the total proportion of the effect of the exposure on the outcome that is mediated by the set. If the intermediate phenotype mediators are correlated, or one also mediates the effect of another on the outcome, the total proportion of the outcome mediated by all of the intermediate phenotypes may be less than the sum of the proportion mediated by each one individually; therefore, each of the above approaches will estimate different effects. A detailed description of the use of MR for mediation analysis is given elsewhere[104].

Non-linear MR. Standard MR provides only a single effect estimate, which may not be informative if the effect of the exposure varies in a non-linear way — for example, a dose–response curve. With individual level data and a continuous exposure, non-linear MR can be applied to estimate whether the causal effect of the exposure on the outcome varies across different levels of the exposure[29,107]. For example, although mortality risk generally increases with BMI, an increase is also seen at very low BMIs; this J-shaped relationship may reflect weight loss in individuals who are unwell, potentially before their illness is diagnosed. Non-linear MR has supported this, although it has also suggested that the J-shape could be caused by the relationship between BMI and mortality risk differing for ever-smokers and never-smokers[108].

Table 2 | **Summary of selected software packages for performing MR analyses**

Package name	Software	Description
Individual-level data		
AER	R	Includes the ivreg function for 2SLS estimation
OneSampleMR	R	Various functions for one-sample IV analyses, including the Sanderson–Windmeijer *F* statistic, and various estimators (two-stage predictor substitution, two-stage residual inclusion, structural mean models)
ivmodel	R	Various functions for individual-level IV analyses, includes LIML, weak instrument tests and sensitivity analyses
ivtools	R	Various functions for individual-level IV analyses, including functions to fit structural mean models
ivonesamplemr	Stata	Includes various estimators (two-stage predictor substitution, two-stage residual inclusion, structural mean models) for one-sample IV analyses
ivreg2	Stata	Stata module for extended IVs/2SLS and generalized method of moments estimation
ivregress	Stata	Linear IV estimators including 2SLS
Summary-level data		
MendelianRandomization	R	Implements several methods for performing MR analyses with summarized data and an interface with the PhenoScanner database
TwoSampleMR and MR-Base app	R/web-app	MR-base is an analytical platform for MR. TwoSampleMR is the R package providing the functions to perform MR estimation. Both are linked to the OpenGWAS project, a large database of GWAS summary statistics
mrrobust	Stata	Provides various programs for two-sample MR analyses in Stata

2SLS, two-stage least-squares; GWAS, genome-wide association study; IV, instrumental variable; LIML, limited information maximum likelihood; MR, Mendelian randomization.

Testing for interactions between exposures. With individual-level data, it is possible to test for interactions between two exposures using MR. When individual-level data are available to conduct a multivariable MR, interactions between the exposures can be included as additional exposures in the estimation[109,110] This requires a multivariable MR estimation including the exposure, the potential effect modifier and the interactions between them included as exposures. The inclusion of these additional terms decreases the statistical power for detecting an effect and should be limited to a single interaction. An alternative approach is to split the allele scores for each exposure into high and low values and to compare outcomes across the resulting four groups by dividing participants up on the basis of their score for each exposure, mimicking a 2 × 2 factorial randomized trial. It should be noted that this approach can have low power compared with the inclusion of an interaction term in a 2SLS regression[110].

Colocalization and MR

Ever larger GWAS have now provided evidence that hundreds of genetic variants can be associated with many human phenotypes. This, together with the tendency for neighbouring genetic variants to be correlated owing to linkage disequilibrium, could lead to the violation of IV condition 2, in which different neighbouring variants happen to be causally associated to the exposure and outcome through different pathways (FIG. 6a). The bias in this situation is equivalent to that caused by pleiotropy (FIG. 3) and, although it is unlikely that this pattern will arise at many independent genetic locations in MR studies with multiple IVs, it should be a consideration in single-IV studies.

Colocalization analysis can be used to determine whether two traits share causal variants in a single genetic region, without prior knowledge of which variant is causal for either trait. It was originally used to identify potential molecular causes of single GWAS associations and considers the patterns of association across multiple neighbouring genetic variants for the GWAS and exposure traits (including molecular traits). Although this involves an implicit assumption of directionality in its interpretation, the test is not dependent on this assumption and indeed a single pleiotropic variant would satisfy the statistical definition of a shared causal variant (FIG. 6b). Unlike in MR with multiple IVs, the majority of multiple neighbouring genetic variants considered in this analysis are expected to be associated with either trait solely through linkage disequilibrium with one or a small number of causal variants in the region. This explicit use of linkage disequilibrium means that colocalization can be used to check for the violation of IV condition 2 in the form shown in FIG. 6a (and FIG. 3c).

One colocalization method originally proposed by Plagnol et al.[111] frames shared causality as the null hypothesis, and rejection of this would indicate violation of IV condition 2, that is, that there are no common causes of the instrument and the outcome[111–113]. However, it is hard to determine whether failure to reject the null hypothesis indicates that IV condition 2 is satisfied or a lack of power in the colocalization test. Alternatively, Bayesian frameworks for colocalization analysis consider GWAS summary statistics for both traits across multiple SNPs in the region around the IV and assess either the evidence that each variant is jointly causal[114] or consider shared causal variants as one of five competing hypotheses[115]. A key difference between MR and Bayesian colocalization strategies is that the latter assume summary data to exist for multiple variants in a region, with sufficient density such that any causal variant or variants for an outcome and exposure are likely to be included in the set of variants studied. This assumption is required because Bayesian colocalization approaches enumerate all possible configurations of causal variants for each trait and assess the relative likelihood of each combination. A further difference is that in Bayesian colocalization strategies the user must supply parameters describing their prior belief that the outcome and exposure share causal variants; these may be different in the context of the carefully chosen traits in MR compared with those in more typical uses of colocalization, and thus sensitivity analyses are recommended to confirm the robustness of inference to changes in prior parameter values[116].

Gene expression and proteins are often instrumented with a single genetic variant and so colocalization can be particularly useful in MR studies of these exposures[41]; in these settings colocalization can be used to attempt to falsify IV condition 2 by testing the hypothesis of a shared causal variant for the exposure and outcome.

Applications

Below we describe five applications of MR. The studies described below have used MR to make important theoretical or practical contributions to understanding the causes of disease and some have implemented recently developed enhanced analytical approaches.

Estimation when trials are unfeasible

Conventional observational epidemiological studies have long suggested a J-shaped relationship between alcohol and risk of cardiovascular disease (CVD)[117–119]. It was unclear from these studies whether the J-shape reflected a true non-linear cause and effect relationship, was caused by confounding by socio-demographic factors, or was present because individuals with low alcohol consumption had a higher apparent risk of CVD owing to a reduction in alcohol consumption caused by sickness (a form of reverse causation known as ‘sick quitters’). Although efforts were made to assess this question using a RCT[120], the trial was terminated by the US National Institutes of Health (NIH) following concerns regarding the study design and influence from the alcohol industry[121–123]. Furthermore, ethical issues exist in deliberately exposing individuals to alcohol, which is a named carcinogen by the International Agency for Research on Cancer (IARC)[124] and is recognized to have multiple detrimental effects on human health including liver disease, depression, and cancers of the oesophagus and liver[125].

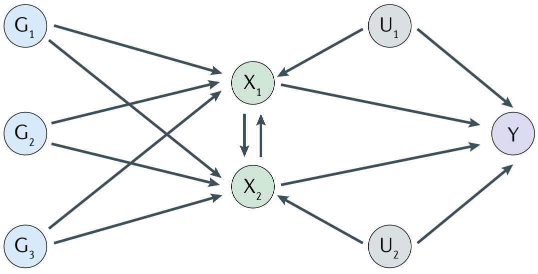


Fig. 5 | **Illustration of the multivariable MR model.** Figure illustrating multivariable Mendelian randomization (MR) for three genetic variants (G_1, G_2, G_3), two exposures (X_1, X_2) and an outcome Y. Confounders U_1 and U_2 are assumed to be unknown.

Early MR studies in individuals with European ancestry using a single genetic variant (rs122994) in the *ADH1B* gene[126,127] suggested that the apparent protective effect of alcohol on the risk of CHD and ischaemic stroke shown in epidemiological studies might not be real. However, use of a single genetic variant with a modest effect on the magnitude of alcohol consumption meant the relationship across the distribution of alcohol consumption could not be explored[128]. In a recent study, Millwood and colleagues[129] used genetic variants in *ALDH2* and *ADH1B*, which together explained considerable variation in alcohol use. Across the distribution of genetic variants, the average amount of alcohol consumed varied from 4 g per week to 256 g per week. Applying these genetic variants to the China Kadoorie Biobank, they found strong evidence of a dose–response relationship between alcohol and risk of stroke, and no strong evidence of a protective or detrimental effect on risk of CHD. In the same study, they were able to show the J-shaped observational association between alcohol and CHD and stroke that had been observed elsewhere. Further, use of negative controls (specifically, exploration of the effect of the genetic instrument in women who did not drink alcohol), empirically demonstrated that the genetic instrument was unlikely to have effects on disease independent of the exposure of interest. Thus, available evidence from MR methods that facilitate estimation in the presence of unobserved confounding — assuming no selection bias — do not support the conclusion that the consumption of a moderate amount of alcohol may lower vascular disease risk and identify alcohol consumption as a factor linked to increased likelihood of ischaemic stroke.

Cholesterol and CHD

Cholesterol circulating in the blood plays a central role in atherosclerosis, the disease process affecting arteries that leads to symptomatic cardiovascular disease including CHD and ischaemic stroke[130]. An inverse association between high-density lipoprotein cholesterol (HDL-C) and CHD risk was reported over a number of observational studies, leading to the widely held belief that high levels of HDL-C are protective against CHD risk[131–134]. This association was observed to be persistent even

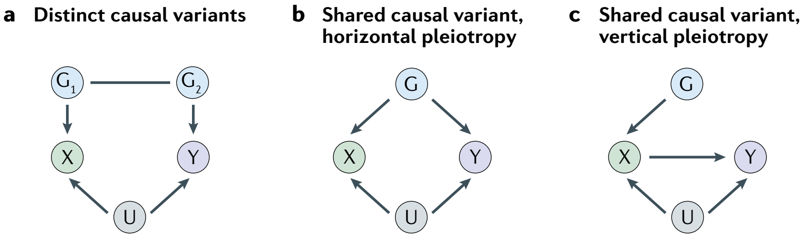


Fig. 6 | **Illustration of variants in linkage disequilibrium and shared causal variants identified by colocalization. a** | An example of distinct causal variants that violate the instrumental variable assumption IV2. G_1 and G_2 represent two genetic variants and the link between them is non-directional, reflecting linkage disequilibrium. **b**,**c** | Examples of a shared causal variant are a violation of assumption IV2 (panel **b**) and a situation that satisfies the IV assumptions (panel **c**).

when other lipid fractions were accounted for, suggesting this association was not owing to confounding[133].

MR studies have provided accumulating evidence against the observational results above[135–138]. Such MR studies used a range of genetic variants that act through different mechanisms and showed no protective effect of increased levels of HDL-C on CHD risk. These studies were published alongside the results of several large-scale RCTs of pharmacological interventions that specifically increased HDL-C without a noticeable change in other blood lipids such as LDL-C, which also failed to show a protective effect[139,140]. These data indicate that the association observed in the more traditional observational studies was likely to have been due to confounding. It is worth reflecting on whether the RCTs would have been embarked upon if the MR study findings were known at the time of their inception[134]. Indeed, where data already exists, MR studies are relatively cheap to conduct — particularly compared with a large RCT — and can provide additional evidence that can be used to direct which studies are worth following up with RCTs. However, it must be noted that MR studies are themselves not free from issues of bias or lack of power; evidence from MR studies for the presence or absence of an effect should be triangulated with findings from studies using different methods that would be expected to have different sources of bias[6,7].

Testing causation across the life course

A key issue in preventing disease in adulthood is identifying when in the life course harmful exposures must be minimized. For example, if the contribution of exposures in childhood is non-reversible, this evidence would argue in favour of early intervention. This is challenging to appraise using conventional observational epidemiology owing to various features such as time-dependent confounding.

One example of this issue is the relationship between adiposity and adult-onset diseases such as CHD and type 2 diabetes (T2D). An MR study[141] took an innovative approach by constructing separate genetic instruments for early-life body size and adult body size. The authors were able to fit a multivariable MR model to elucidate whether childhood body size was detrimental to the risk of CHD or T2D after taking adult body size into account. A direct effect of childhood body size in the multivariable model would suggest that high adiposity in childhood has a long-term effect on health outcomes in adulthood — suggesting that focusing on early interventions in childhood to minimize excess body weight would be helpful in lowering the risk of diseases that typically present in adulthood. As UK Biobank participants were asked for information on their body size at 10 years of age and BMI was measured at recruitment into the study[142], these data provided an opportunity to conduct GWAS on body size during childhood and adulthood for the same group of individuals and detected 295 and 557 independent SNPs associated with childhood and adulthood body size, respectively, with a high level of overlap in the SNPs associated with each time period, as expected[141]. Univariable MR analysis showed that both genetically predicted body size in early life and adulthood were individually related to higher risks of CHD and T2D and a lower risk of breast cancer. By contrast, multivariable MR analysis identified that only adult body size showed an independent causal effect for CHD and T2D, suggesting that the relationship between early-life body size was mediated through adult body size. By contrast, the inverse relationship between genetically predicted body size and breast cancer was stronger for early-life body size than adult body size in the multivariable MR analysis, suggesting an age-dependent relationship between adiposity and the risk of different diseases in adults. This suggests that for children that are overweight, losing weight in their adulthood can still effectively lower risk of T2D and CAD and in this case a metabolically unhealthy childhood can potentially be offset by healthy lifestyle approaches adopted in adulthood.

Such study designs can be applied to other exposure–outcome relationships to determine whether risk factors have cumulative effects or differential influences at different periods of the life course. This information could allow for fine-tuned, age-specific public health interventions that minimize the effects of deleterious, time-dependent risk factors. However, it is very important to bear in mind that effects of harmful exposures may become less evident with increasing age because of selection bias owing to the almost inevitable selection of survivors[143].

Estimation of health-care costs

A clear understanding of the health-care costs arising from individual diseases and risk factors is needed to ensure that public health resources are distributed judiciously. RCTs are typically not designed to estimate health-care costs as an outcome and conventional observational studies aimed at assessing health-care costs can be hampered by selection bias and confounding.

Dixon and colleagues[144] described a potential application for MR in quantifying the effects of genetically predicted BMI on health-care costs. Their method used data from the UK Biobank, which provided a rich source of data for exploring the causal relationship of lifelong exposures to certain traits and genetic liability to diseases and their economic impact. Using genetic variants associated with higher BMI as instruments in an individual-level MR study to estimate the effect of BMI on hospitalization costs[145], the authors found that higher BMI increased hospital costs with little evidence for non-linearity in this effect. In addition to physiological consequences, body weight has social consequences such as increasing exposure to stigma and discrimination and these MR analyses include the consequences of all such mechanisms for hospitalization costs.

Testing treatment response factors

Identifying whether individuals are likely to respond to a specific therapy is an important component of so-called 'precision medicine', whereby the goal is to individualize patient care based on genetic, environmental and lifestyle factors. This can be done in conventional pharmacogenetic studies and RCTs, although the risk of bias in the former and the sample size constraints of the latter mean that neither provide a reliable means of assessing interactions between an individual's genotype and treatment response.

A recent study by Xu and Burgess[146] used a drug-target MR design[41,147] to investigate polygenic determinants of the response of LDL-cholesterol levels to treatment with statins. The authors used SNPs in and around the *HMGCR* locus as a mimic of the pharmacological inhibition of HMG-CoA reductase by statins, and explored genetic variants that might act as effect modifiers of the association between the statin genetic instrument and LDL-cholesterol levels. Polygenic scores did not identify any effect-modifying genetic groups; however, a single variant (rs162724) proximal to the glutamate receptor gene *GRM7* and previously associated with major depressive disorder was found to potentially be of interest. The authors postulated that this variant could be related to statin response via concurrent pharmacotherapies for major depressive disorder or via poorer adherence to statin treatment reducing the effect of statins on LDL-cholesterol.

Although the above study did not find evidence of reliable polygenic effect modification, it introduces the concept of agnostic identification of pharmacogenetic interactions within the context of a population-based study. This approach benefits from lack of confounding by indication, compared with a conventional pharmacoepidemiology study design[148]. However, using a genetic instrument for treatment as part of a drug-target MR means that the underlying magnitude of the effect for which potential genetic effect modifiers are investigated is very small and thus very large sample sizes are needed to identify effects. When using MR in this way, it is important to identify appropriate instruments for estimating the effect of a particular drug. Instruments that are associated with the target of that drug should be used, rather than those associated with the risk factor that the drug acts on[41,149,150].

Reproducibility and data deposition

There has been substantial discussion of the importance of ensuring that published research findings are robust, replicable and reproducible in recent years[151]. In the context of epidemiological research, one area of concern is that findings may be replicated in settings with nearly identical sources of bias. Data with such replication provide little independent confirmation of the initial result and thus even highly consistent replicated findings may not reflect true causal effects. An example is the J-shaped association between alcohol consumption and cardiovascular disease; there is now consensus that this apparent protective effect of moderate levels of consumption is artefactual, as discussed above[129]. One simple step authors can take to ensure that MR findings are robust and reproducible is to use the STROBE-MR guidelines[50,51], which outline how MR studies should be reported to make the approach used in any particular study clear for readers.

The first aim of all studies should be to ensure that steps are taken to detect and minimize biases such as selection bias or bias caused by violation of one of the IV conditions. Triangulation of evidence from multiple methodologies — using different methodologies that are subject to different sources and directions of potential bias — can help to identify bias in MR studies[6,7,152]. Alignment of results across these different methodologies can improve confidence in an initial causal interpretation. Among the most promising strategies for triangulation is contrasting MR results with results using other IVs — such as policy-based IVs — or results from conventional analyses. For example, there is clear evidence from both MR and the natural experiment of an increase in the school leaving age that an increase in the number of years in education has a causal protective effect on health behaviours such as smoking[153–156] (see preprint[157]). Within MR, using methods that make different assumptions (such as those regarding pleiotropy) and are therefore subject to different sources and directions of potential bias can support this approach, although some important assumptions may be shared by many methods, reducing the potential independent insight to be gained from comparing studies.

Open research can increase the robustness of data by enabling greater scrutiny of data and increased error detection by researchers and the wider research community. Open research approaches for increasing data transparency include protocol pre-registration and sharing of data, code and materials. Summary data from GWAS are often a source of data for MR analysis and are typically publicly available, such as those listed on the OpenGWAS project. Although individual-level data are not made publicly available owing to the sensitive nature of the data, there are a number of large datasets that are accessible to any researchers on application, such as the UK Biobank. Any MR estimation should clearly indicate the data sources they have used and link to the dataset used if it is publicly available. The source code for many software packages is openly available (for example, TwoSampleMR and mrrobust on GitHub, and MendelianRandomization on CRAN). Although the analysis code from MR studies is not routinely shared, we would encourage readers of this Primer to do so to enable errors in coding to be more readily identified. Pre-registration of study protocols has not been widely adopted in observational epidemiology, although it could in principle be applied and would help to protect against biases such as publication bias against null results or findings that do not fit with the anticipated conclusion[158] (see preprint[159]).

Limitations and optimizations

An important limitation of MR studies is the potential confounding of the genetic variants and the outcome (violation of IV condition 2; BOX 2). As genetic variants are generally fixed at conception, it is not intuitively clear how confounding of the instrument and the outcome can occur in MR studies. However, population stratification, dynastic effects and assortative mating all induce bias by creating an artefactual relation at the population level between the genetic variants and the outcome, violating the second IV condition[64,160–163]. Each of these sources of confounding are described in detail in BOX 5. This correlation between genetic variants and the outcome can potentially affect most (or all) of the genetic variants used as instruments; it is therefore not easy to correct for using current MR methods given that most assume that the majority of genetic variants satisfy all of the IV conditions[60]. Considering the potential for biases of the sort described here is therefore crucial in the interpretation of any MR result.

Box 5 | **Sources of instrument–outcome confounding in MR studies**

Population stratification

Population stratification is the association between genetic variants and phenotypes that occurs because of underlying structure within the population[52,220]. This underlying structure reflects the fact that genetic mutations accrue and accumulate across generations, and that individuals differentially select partners who are geographically proximal. Within genome-wide association studies, population stratification is often controlled for by adjusting for the top principal components from a principal component analysis of the genetic variants or through the use of linear mixed models[221–223]. However, there is increasing evidence that these approaches do not fully account for the underlying structure for a number of phenotypes[224,225]. Population stratification can bias estimates from Mendelian randomization (MR) studies by creating an association between the genetic variants and the outcome as illustrated in panel **a** of the figure[161,224]. In the figure, G represents genetic variants, X represents exposure and Y represents outcome in a MR study.

Dynastic effects

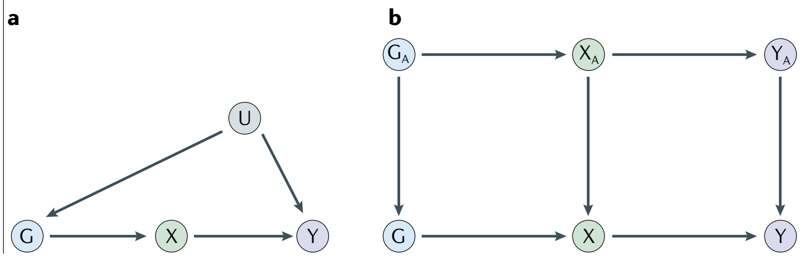
Dynastic effects are the direct effects on an individual's phenotypes of the phenotypes of their parents, and (potentially to a lesser extent) more distantly related relatives such as grandparents. As parental genotypes have a direct effect on the genotype of an individual, if a parent's phenotype is influenced by their genotype and influences the individual's phenotype this will induce confounding between the genetic variants and phenotype of the offspring, as illustrated in panel **b** of the figure[162]. If the exposure has a non-null causal effect on the outcome in a MR study, these dynastic effects will induce instrument–outcome confounding and bias the results of the MR study[161]. In the figure, G_A, X_A and Y_A are the genetic variants, exposure and outcome respectively for ancestors (such as parents) of the individuals under consideration in the MR estimation.

Assortative mating

Assortative mating occurs when individuals select partners who are more similar to themselves than would be expected by chance, with respect to one or multiple phenotypes[226,227]. If the genetically influenced level of the phenotype influences selection, this assortment can lead to spurious genetic associations with the phenotype or phenotypes on which the assortment is based or that are causally dependent on the assortment phenotypes. This consequently biases MR estimates involving these phenotypes[160,161].

Transmission ratio distortion

Transmission ratio distortion occurs when the transmission of alleles from parents to offspring deviates from the expected probability of 50:50. This can occur owing to processes during meiosis and fertilization favouring one parental allele over another or if the viability of the offspring depends on their genotype. If environmental factors influence the transmission ratio distortion, those environmental factors will become associated with genotype in the offspring. The association between any environmental factor and the genotype can lead to the potential for instrument–outcome confounding in MR if the environmental factor also influences the outcome[164]. Until recently, data on parent–offspring trios were not available at the scale required to investigate this possibility, but this is now becoming possible[228].



One solution that can account for confounding owing to dynastic effects and assortative mating is the use of family data to conduct the MR analyses[164,165]. Within-family MR requires data from either pairs of siblings or mother–father–child trios and allows for the estimation of causal effects using MR after family-level structure has been taken into account[161,164]. Within-family MR using sibling pairs will also account for any factors acting at a population level that affect siblings equally, such as population stratification. A key limiting factor for within-family MR is the lack of available data and the low power of these studies as a result; however, a GWAS of family data for a range of phenotypes has recently been published, enabling further within-family MR in the future[166].

Another type of bias that can arise in MR studies that cannot be easily corrected for is selection bias[63]. In an MR study, an example of selection bias would be if an individual's exposure and outcome values affected their participation[56]. When these phenotypes are partially determined by genetic variants, this will also induce an association between those genetic variants and participation. Study participation has been shown to be heritable and is influenced by a number of different traits, and large studies such as the UK Biobank have been shown to have high levels of selection in those who participate[30,167–169].

In addition, most studies recruit survivors of the original birth cohorts. This means all participants must have survived in order to observe whether they get the outcome of interest. Selection of participants on surviving their genetic make-up and the outcome of interest or a competing risk of the outcome effectively applies covariable adjustment on survival into the estimates[170–172]. This form of selection bias is likely to be particularly problematic for studies of harmful exposures on disease outcomes that occur in later life and will be least evident in studies where the exposure does not affect survival to recruitment[173]. As such, consideration of whether the genetically instrumented exposures would affect survival to recruitment, age at recruitment or any competing risk of the outcome may help to identify bias. This type of survival bias will affect observational studies of the same research question in similarly aged populations, so it is not an obvious explanation for discrepancies between MR and conventional results. All forms of selection bias could bias MR estimates and so careful assessment of the potential for selection into the sample or samples used in an MR study is important[64]. Novel methods are being developed that attempt to detect and correct for selection bias[171,174]; however, this is an area in which further research is required.

Finally, MR uses genetic variants that are fixed across the life course to estimate the lifetime effects of the exposure of interest. This introduces a potential limitation in the form of canalization, which refers to a natural tendency for the suppression of phenotypic variation among individuals despite contrasting genotypes. Canalization can occur when polymorphic phenotypes expressed during fetal development lead to the development of compensating pathways to mitigate the effects of that expression[1,175,176]. For example, individuals with genetically elevated fibrinogen levels could become resistant to the effects of higher fibrinogen owing to permanent changes in tissue structure during fetal development. Canalization is seen following dramatic genetic or environmental changes, for example in gene-knockout studies[177,178]. Such compensation would potentially limit the ability of MR to identify the causal effect of the change in the exposure because the effect of a genetically induced change from conception would be different from the effect of a change in later life. This is an example of a violation of the assumption of gene–environment equivalence (BOX 4). Further work is required to understand whether small changes induced by the common polymorphisms used to estimate causal effects in MR lead to the same compensatory effects.

Outlook

The rapid increase in MR publications demonstrates the need for approaches that can contribute to strengthening causal inference. This growth in the quantity of published MR studies comes with anxiety regarding their quality. Papers reporting two-sample MR have grown rapidly over recent years and now constitute a large majority of published studies[8,80]. These are relatively easy to conduct — perhaps too easy — and they can contain obvious errors, as discussed and demonstrated elsewhere[80]. Indeed, many such papers simply report MR estimates obtained from applying open-access software to open-access data and in these cases the analyses have, in essence, already been conducted by automated tools — an observation detailed in a preprint article[179]. The situation with MR is now moving towards the one seen in the meta-analysis literature, with the mass production of redundant, misleading and conflicted publications[180]. The current explosion in predatory journals unfortunately means that this situation is very unlikely to change. There are now a number of guidelines available for MR estimation, and those regarding the conduct[39] and reporting of MR studies[50,51] are useful for understanding and identifying whether a MR study has been well conducted and reported properly. For those aiming to keep up with the MR literature, the twitter account @MR_lit searches for papers and preprint articles and allows readers to rapidly review abstracts to identify papers of interest.

As most contemporary MR studies rely on available GWAS data, they unfortunately suffer from considerable bias with respect to the representativeness of populations according to geography and ancestry[181]. This can influence the generalizability of MR findings and exacerbate existing inequity in medical research. It can also restrict the scope of MR studies, as some forms of genetic variation are restricted to particular populations. For example, a large-effect genetic variant influencing alcohol consumption that has been of considerable value in MR studies of the effects of alcohol[74,129] is only prevalent in East Asian populations. Current international efforts to equalize inclusion of different populations in genetic studies will hopefully begin to address this important issue.

A large area of medical research is aimed at identifying potentially therapeutic influences on disease progression once the disease is established. However, MR studies usually rely on GWAS of the initial development of disease for their outcome data. This means that although MR has been a powerful tool for confirming or discovering factors that cause disease, it does not often identify therapeutic targets[182]. For example, although MR studies have shown that smoking causes lung cancer[183], this is not useful therapeutically following the onset of the disease given that smoking cessation is not a useful treatment once lung cancer has developed. It is plausible that in many cases, factors that cause a disease do not relate to its progression once it is established. For example, the onset and progression of Crohn's disease are associated with different genetic variants, indicating that different risk factors play a part in onset and development[184]. It is also possible that the same risk factor could have opposite effects on incidence and progression, as has been suggested for folate intake and colon cancer[185]. MR of factors influencing disease progression is needed to identify useful treatments[186]; however, such estimation requires appropriate datasets and as there are currently few of these in existence, efforts should be focused on increasing the availability of such data. Importantly, case-only study designs may be severely compromised by collider bias[55,63], which must be taken into account in data analysis[182]. Further methodological development is required in this domain.

Although the increasing size of GWAS datasets appears to be good news for MR studies, it can also introduce problems; smaller and smaller effect sizes are being identified as 'genome-wide significant' in GWAS and it is increasingly likely that such variants affect the trait of interest through an upstream phenotype that might in turn influence the outcomes under investigation. For example, as the GWAS of CRP and vitamin D increased in size, multiple variants that primarily influence adiposity were identified — with adiposity being a confounder of the observational association of these exposures with health outcomes. If these variants are used as instruments for CRP or vitamin D, they will produce highly misleading results. The resulting bias can be accounted for through multivariable MR if the upstream factor is known; however, in many cases it is not known and so the bias will remain undetected. This issue of misspecification of the primary phenotype[8,61] requires

Collider bias
Bias occurring owing to conditioning on a variable that is dependent on both the exposure and outcome or is dependent on causes of the exposure and outcome.

more research to identify the extent of the problem of recapitulating confounding in MR studies as GWAS size increases.

When initially presented, it was concluded that "[MR] offers a more robust approach to understanding the effect of some modifiable exposures on health outcomes than does much conventional observational epidemiology"[1] and that, where possible, RCTs should follow to establish the effects of interventions. This conclusion remains unchanged, although moving towards formal triangulation[6,7,152] of all pertinent evidence as discussed in this Primer should be the goal of all research aimed at identifying causal influences on health and development outcomes.

1. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
2. Angrist, J. D., Imbens, G. W. & Rubin, D. B. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* **91**, 444–455 (1996).
3. Hernán, M. A. & Robins, J. M. *Causal Inference: What If* (Chapman & Hall/CRC, 2020).
4. Greenland, S. An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.* **29**, 722–729 (2000).
5. Zuccolo, L. & Holmes, M. V. Commentary: Mendelian randomization-inspired causal inference in the absence of genetic data. *Int. J. Epidemiol.* **46**, 962–965 (2017).
6. Munafò, M. R., Higgins, J. P. & Davey Smith, G. Triangulating evidence through the inclusion of genetically informed designs. *Cold Spring Harb. Perspect. Med.* **11**, a040659 (2021).
7. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2017).
8. Richmond, R. C. & Davey Smith, G. Mendelian randomization: concepts and scope. *Cold Spring Harb. Perspect. Med.* https://doi.org/10.1101/cshperspect.a040501 (2022).
9. Davey Smith, G. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *Int. J. Epidemiol.* **33**, 30–42 (2004).
10. Gupta, S. K. Intention-to-treat concept: a review. *Perspect. Clin. Res.* **2**, 109–112 (2011).
11. Ellenberg, J. H. Intent-to-treat analysis versus as-treated analysis. *Drug Inf. J.* **30**, 535–544 (1996).
12. Glymour, M. M. Natural experiments and instrumental variable analyses in social epidemiology. *Methods Soc. Epidemiol.* **1**, 429 (2006).
13. Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V. & Klungel, O. H. Instrumental variables: application and limitations. *Epidemiology* **17**, 260–267 (2006).
14. Lousdal, M. L. An introduction to instrumental variable assumptions, validation and estimation. *Emerg. Themes Epidemiol.* **15**, 1 (2018).
15. Angrist, J. D. & Krueger, A. B. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J. Econ. Perspect.* **15**, 69–85 (2001).
16. Rassen, J. A., Brookhart, M. A., Glynn, R. J., Mittleman, M. A. & Schneeweiss, S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J. Clin. Epidemiol.* **62**, 1226–1232 (2009).
17. Didelez, V. & Sheehan, N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat. Methods Med. Res.* **16**, 309–330 (2007).
18. Davey Smith, G. Capitalizing on Mendelian randomization to assess the effects of treatments. *J. R. Soc. Med.* **100**, 432–435 (2007).
19. Carlson, C. S. et al. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *Am. J. Hum. Genet.* **77**, 64–77 (2005).
20. Davey Smith, G. et al. Association of C-reactive protein with blood pressure and hypertension: life course confounding and Mendelian randomization tests of causality. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1051–1056 (2005).
21. Morris, T. T., Heron, J., Sanderson, E., Davey Smith, G. & Tilling, K. Interpretation of Mendelian randomization using one measure of an exposure that varies over time. Preprint at *medRxiv* https://doi.org/10.1101/2021.11.18.21266515 (2021).
22. Swanson, S. A., Tiemeier, H., Ikram, M. A. & Hernán, M. A. Nature as a trialist? Deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiology* **28**, 653–659 (2017).
23. Didelez, V., Meng, S. & Sheehan, N. A. Assumptions of IV methods for observational epidemiology. *Statist. Sci.* **25**, 22–40 (2010).
24. Palmer, T. M. et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat. Methods Med. Res.* **21**, 223–242 (2011).
25. Burgess, S. & Thompson, S. G. Use of allele scores as instrumental variables for Mendelian randomization. *Int. J. Epidemiol.* **42**, 1134–1144 (2013).
26. Davies, N. M. et al. The many weak instruments problem and Mendelian randomization. *Stat. Med.* **34**, 454–468 (2015).
27. Hernán, M. A. & Robins, J. M. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* **17**, 360–372 (2006).
28. Swanson, S. A., Hernán, M. A., Miller, M., Robins, J. M. & Richardson, T. S. Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes. *J. Am. Stat. Assoc.* **113**, 933–947 (2018).
29. Staley, J. R. & Burgess, S. Semiparametric methods for estimation of a nonlinear exposure–outcome relationship using instrumental variables with application to Mendelian randomization. *Genet. Epidemiol.* **41**, 341–352 (2017).
30. Tyrrell, J. et al. Genetic predictors of participation in optional components of UK Biobank. *Nat. Commun.* **12**, 886 (2021).
31. Davey Smith, G. Epigenesis for epidemiologists: does evo-devo have implications for population health research and practice? *Int. J. Epidemiol.* **41**, 236–247 (2012).
32. Freeman, G., Cowling, B. J. & Schooling, C. M. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *Int. J. Epidemiol.* **42**, 1157–1163 (2013).
33. Walker, V. M., Davies, N. M., Windmeijer, F., Burgess, S. & Martin, R. M. Power calculator for instrumental variable analysis in pharmacoepidemiology. *Int. J. Epidemiol.* **46**, 1627–1632 (2017).
34. Burgess, S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int. J. Epidemiol.* **43**, 922–929 (2014).
35. Brion, M.-J. A., Shakhbazov, K. & Visscher, P. M. Calculating statistical power in Mendelian randomization studies. *Int. J. Epidemiol.* **42**, 1497–1501 (2012).
36. Morris, T. P., White, I. R. & Crowther, M. J. Using simulation studies to evaluate statistical methods. *Stat. Med.* **38**, 2074–2102 (2019).
37. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
38. Zhao, Q., Wang, J., Spiller, W., Bowden, J. & Small, D. S. Two-sample instrumental variable analyses using heterogeneous samples. *Stat. Sci.* **34**, 317–333 (2019).
39. Burgess, S. et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res.* **4**, 186 (2019).
40. Pierce, B. L. & Burgess, S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am. J. Epidemiol.* **178**, 1177–1184 (2013).
41. Holmes, M. V., Richardson, T. G., Ference, B. A., Davies, N. M. & Davey Smith, G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nat. Rev. Cardiol.* **18**, 435–453 (2021).
42. Bound, J., Jaeger, D. A. & Baker, R. M. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J. Am. Stat. Assoc.* **90**, 443–450 (1995).
43. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
44. Mounier, N. & Kutalik, Z. Correction for sample overlap, winner's curse and weak instrument bias in two-sample Mendelian Randomization. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.26.437168 (2021).
45. Angrist, J. D. & Krueger, A. B. Split-sample instrumental variables estimates of the return to schooling. *J. Bus. Econ. Stat.* **13**, 225–235 (1995).
46. Fang, S., Hemani, G., Richardson, T. G., Gaunt, T. R. & Davey Smith, G. Evaluating and implementing block jackknife resampling Mendelian randomization to mitigate bias induced by overlapping samples. Preprint at *medRxiv* https://doi.org/10.1101/2021.12.03.21267246 (2021).
47. Sadreev, I. I. et al. Navigating sample overlap, winner's curse and weak instrument bias in Mendelian randomization studies using the UK Biobank. Preprint at *medRxiv* https://doi.org/10.1101/2021.06.28.21259622 (2021).
48. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).
49. Holmes, M. V., Ala-Korpela, M. & Davey Smith, G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat. Rev. Cardiol.* **14**, 577–590 (2017).
50. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* **375**, n2233 (2021).
51. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *JAMA* **326**, 1614–1621 (2021).
52. Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
53. Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data* (MIT Press, 2010).
54. Cole, S. R. et al. Illustrating bias due to conditioning on a collider. *Int. J. Epidemiol.* **39**, 417–420 (2009).
55. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* **47**, 226–235 (2018).
56. Hernán, M. A., Hernández-Díaz, S. & Robins, J. M. A structural approach to selection bias. *Epidemiology* **15**, 615–625 (2004).
57. Staiger, D. & Stock, J. H. Instrumental variables regression with weak instruments. Report No. 0898-2937 (National Bureau of Economic Research, 1994).
58. Stock, J. H. & Yogo, M. Testing for weak instruments in linear IV regression. Report No. 0898-2937 (National Bureau of Economic Research, 2002).
59. Brumpton, B. et al. Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. *Nat. Commun.* **11**, 1–13 (2020).
60. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27**, R195–R208 (2018).
61. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–R98 (2014).
62. Burgess, S., Swanson, S. A. & Labrecque, J. A. Are Mendelian randomization investigations immune from bias due to reverse causation? *Eur. J. Epidemiol.* **36**, 253–257 (2021).

63. Griffith, G. J. et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat. Commun.* **11**, 5749 (2020).
64. Hughes, R. A., Davies, N. M., Davey Smith, G. & Tilling, K. Selection bias when estimating average treatment effects using one-sample instrumental variable analysis. *Epidemiology* **30**, 350–357 (2019).
65. Sargan, J. D. The estimation of economic relationships using instrumental variables. *Econometrica* **26**, 393–415 (1958).
66. Glymour, M. M., Tchetgen Tchetgen, E. J. & Robins, J. M. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am. J. Epidemiol.* **175**, 332–339 (2012).
67. Diemer, E. W., Labrecque, J., Tiemeier, H. & Swanson, S. A. Application of the instrumental inequalities to a Mendelian randomization study with multiple proposed instruments. *Epidemiology* **31**, 65–74 (2020).
68. Yang, Q., Sanderson, E., Tilling, K., Borges, M. C. & Lawlor, D. A. Exploring and mitigating potential bias when genetic instrumental variables are associated with multiple non-exposure traits in Mendelian randomization. Preprint at *medRxiv* https://doi.org/10.1101/19009605 (2019).
69. Lawlor, D. A. et al. Exploring the developmental overnutrition hypothesis using parental–offspring associations and FTO as an instrumental variable. *PLoS Med.* **5**, e33 (2008).
70. Kang, H., Zhang, A., Cai, T. T. & Small, D. S. Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *J. Am. Stat. Assoc.* **111**, 132–144 (2016).
71. Windmeijer, F., Farbmacher, H., Davies, N. & Davey Smith, G. On the use of the lasso for instrumental variables estimation with some invalid instruments. *J. Am. Stat. Assoc.* **114**, 1339–1350 (2019).
72. Jiang, L. et al. Constrained instruments and their application to Mendelian randomization with pleiotropy. *Genet. Epidemiol.* **43**, 373–401 (2019).
73. 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.* **48**, 713–727 (2019).
74. Chen, L., Davey Smith, G., Harbord, R. M. & Lewis, S. J. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLoS Med.* **5**, e52 (2008).
75. Spiller, W., Hartwig, F. P., Sanderson, E., Davey Smith, G. & Bowden, J. Interaction-based Mendelian randomization with measured and unmeasured gene-by-covariate interactions. Preprint at *medRxiv* https://doi.org/10.1101/2020.07.27.20162909 (2020).
76. Spiller, W., Slichter, D., Bowden, J. & Davey Smith, G. Detecting and correcting for bias in Mendelian randomization analyses using gene-by-environment interactions. *Int. J. Epidemiol.* **48**, 702–712 (2019).
77. Tchetgen Tchetgen, E. J., Sun, B. & Walter, S. The GENIUS approach to robust Mendelian randomization inference. *Stat. Sci.* **36**, 443–464 (2019).
78. Burgess, S., Dudbridge, F. & Thompson, S. G. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat. Med.* **35**, 1880–1906 (2016).
79. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
80. Hartwig, F. P., Davies, N. M., Hemani, G. & Davey Smith, G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* **45**, 1717–1726 (2017).
81. Bowden, J. 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.* **47**, 1264–1278 (2018).
82. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
83. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
84. Rees, J. M., Wood, A. M., Dudbridge, F. & Burgess, S. Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. *PloS ONE* **14**, e0222362 (2019).
85. Cho, Y. et al. Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework. *Nat. Commun.* **11**, 1010 (2020).
86. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
87. Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D. S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann. Stat.* **48**, 1742–1769 (2020).
88. Wang, J. et al. Causal inference for heritable phenotypic risk factors using heterogeneous genetic instruments. *PLoS Genet.* **17**, e1009575 (2021).
89. Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M. & He, X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat. Genet.* **52**, 740–747 (2020).
90. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
91. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).
92. Bowden, J. & Vansteelandt, S. Mendelian randomization analysis of case-control data using structural mean models. *Stat. Med.* **30**, 678–694 (2011).
93. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
94. Brown, B. C. & Knowles, D. A. Welch-weighted Egger regression reduces false positives due to correlated pleiotropy in Mendelian randomization. *Am. J. Hum. Genet.* **108**, 2319–2335 (2021).
95. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018).
96. Elsworth, B. L. et al. The MRC IEU OpenGWAS data infrastructure. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.10.244293 (2020).
97. Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* **396**, 413–446 (2020).
98. Snoeckx, R. L. et al. GJB2 mutations and degree of hearing loss: a multicenter study. *Am. J. Hum. Genet.* **77**, 945–957 (2005).
99. Hoffmann, T. J. et al. A large genome-wide association study of age-related hearing impairment using electronic health records. *PLoS Genet.* **12**, e1006371 (2016).
100. Lambert, J.-C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
101. Burgess, S., Dudbridge, F. & Thompson, S. G. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". *Am. J. Epidemiol.* **181**, 290–291 (2015).
102. Zuber, V., Colijn, J. M., Klaver, C. & Burgess, S. Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nat. Commun.* **11**, 29 (2020).
103. Sanderson, E. Multivariable Mendelian randomization and mediation. *Cold Spring Harb. Perspect. Med.* **11**, a038984 (2020).
104. Carter, A. R. et al. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *Eur. J. Epidemiol.* **36**, 465–478 (2021).
105. Relton, C. L. & Davey Smith, G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int. J. Epidemiol.* **41**, 161–176 (2012).
106. Burgess, S., Daniel, R. M., Butterworth, A. S., Thompson, S. G. & Consortium, E.-I. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int. J. Epidemiol.* **44**, 484–495 (2015).
107. Burgess, S., Davies, N. M. & Thompson, S. G. Instrumental variable analysis with a nonlinear exposure–outcome relationship. *Epidemiology* **25**, 877 (2014).
108. Sun, Y.-Q. et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. *BMJ* **364**, l1042 (2019).
109. North, T.-L. et al. Using genetic instruments to estimate interactions in Mendelian randomization studies. *Epidemiology* **30**, e33–e35 (2019).
110. Rees, J., Foley, C. N. & Burgess, S. Factorial Mendelian randomization: using genetic variants to assess interactions. *Int. J. Epidemiol.* **49**, 1147–1158 (2019).
111. Plagnol, V., Smyth, D. J., Todd, J. A. & Clayton, D. G. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. *Biostatistics* **10**, 327–334 (2009).
112. Wallace, C. Statistical testing of shared genetic control for potentially related traits. *Genet. Epidemiol.* **37**, 802–813 (2013).
113. Pavlides, J. M. W. et al. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med.* **8**, 84–84 (2016).
114. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
115. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
116. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLoS Genet.* **16**, e1008720 (2020).
117. Marmot, M. & Brunner, E. Alcohol and cardiovascular disease: the status of the U shaped curve. *BMJ* **303**, 565–568 (1991).
118. Corrao, G., Rubbiati, L., Bagnardi, V., Zambon, A. & Poikolainen, K. Alcohol and coronary heart disease: a meta-analysis. *Addiction* **95**, 1505–1523 (2000).
119. Mukamal, K. J. & Rimm, E. B. Alcohol's effects on the risk for coronary heart disease. *Alcohol. Res. Health* **25**, 255–261 (2001).
120. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT03169530 (2019).
121. Dyer, O. $100m alcohol study is cancelled amid pro-industry "bias". *BMJ* **361**, k2689 (2018).
122. Mitchell, G., Lesch, M. & McCambridge, J. Alcohol industry involvement in the moderate alcohol and cardiovascular health trial. *Am. J. Public Health* **110**, 485–488 (2020).
123. National Institutes of Health. NIH to end funding for Moderate Alcohol and Cardiovascular Health trial. *National Institutes of Health* https://www.nih.gov/news-events/news-releases/nih-end-funding-moderate-alcohol-cardiovascular-health-trial (2018).
124. Wild, C. in *World Cancer Report 2014* (eds Wild, C. P. & Stewart, B. W.) (World Health Organization, 2014).
125. Secretan, B. et al. A review of human carcinogens — Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *Lancet Oncol.* **10**, 1033–1034 (2009).
126. Lawlor, D. A. et al. Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian randomization study in the Copenhagen General Population Study. *Eur. Heart J.* **34**, 2519–2528 (2013).
127. Holmes, M. V. et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ* **349**, g4164 (2014).
128. Silverwood, R. J. et al. Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. *Int. J. Epidemiol.* **43**, 1781–1790 (2014).
129. Millwood, I. Y. et al. Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. *Lancet* **393**, 1831–1842 (2019).
130. Goldstein, J. L. & Brown, M. S. A century of cholesterol and coronaries: from plaques to genes to statins. *Cell* **161**, 161–172 (2015).
131. Miller, G. & Miller, N. Plasma-high-density-lipoprotein concentration and development of ischaemic heart-disease. *Lancet* **305**, 16–19 (1975).

132. Castelli, W. P. et al. HDL cholesterol and other lipids in coronary heart disease. The cooperative lipoprotein phenotyping study. *Circulation* **55**, 767–772 (1977)
133. Emerging Risk Factors Collaboration et al. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* **302**, 1993–2000 (2009).
134. Davey Smith, G. & Phillips, A. N. Correlation without a cause: an epidemiological odyssey. *Int. J. Epidemiol.* **49**, 4–14 (2020).
135. Voight, B. F. et al. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* **380**, 572–580 (2012).
136. Do, R. et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **45**, 1345–1352 (2013).
137. Holmes, M. V. et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur. Heart J.* **36**, 539–550 (2015).
138. Holmes, M. V. & Davey Smith, G. REVEALing the effect of CETP inhibition in cardiovascular disease. *Nat. Rev. Cardiol.* **14**, 635–636 (2017).
139. Barter, P. J. et al. Effects of torcetrapib in patients at high risk for coronary events. *N. Engl. J. Med.* **357**, 2109–2122 (2007).
140. Riaz, H. et al. Effects of high-density lipoprotein targeting treatments on cardiovascular outcomes: a systematic review and meta-analysis. *Eur. J. Prev. Cardiol.* **26**, 533–543 (2019).
141. Richardson, T. G., Sanderson, E., Elsworth, B., Tilling, K. & Davey Smith, G. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: Mendelian randomisation study. *BMJ* **369**, m1203 (2020).
142. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
143. Schooling, C. M. Selection bias in population-representative studies? A commentary on Deaton and Cartwright. *Soc. Sci. Med.* **210**, 70 (2018).
144. Dixon, P., Davey Smith, G., von Hinke, S., Davies, N. M. & Hollingworth, W. Estimating marginal healthcare costs using genetic variants as instrumental variables: Mendelian randomization in economic evaluation. *PharmacoEconomics* **34**, 1075–1086 (2016).
145. Dixon, P., Hollingworth, W., Harrison, S., Davies, N. M. & Davey Smith, G. Mendelian randomization analysis of the causal effect of adiposity on hospital costs. *J. Health Econ.* **70**, 102300 (2020).
146. Xu, Z. M. & Burgess, S. Polygenic modelling of treatment effect heterogeneity. *Genet. Epidemiol.* **44**, 868–879 (2020).
147. Holmes, M. V. Human genetics and drug development. *N. Engl. J. Med.* **380**, 1076–1079 (2019).
148. Kyriacou, D. N. & Lewis, R. J. Confounding by indication in clinical research. *JAMA* **316**, 1818–1819 (2016).
149. Schmidt, A. F. et al. Genetic drug target validation using Mendelian randomisation. *Nat. Commun.* **11**, 3255 (2020).
150. Schmidt, A. F., Hingorani, A. D. & Finan, C. Human genomics and drug development. *Cold Spring Harb. Perspect. Med*. https://doi.org/10.1101/cshperspect.a039230 (2021).
151. Munafò, M. R. et al. A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
152. Munafò, M. R. & Davey Smith, G. Robust research needs many lines of evidence. *Nature* **553**, 399–401 (2018).
153. Davies, N. M., Dickson, M., Davey Smith, G., van den Berg, G. J. & Windmeijer, F. The causal effects of education on health outcomes in the UK Biobank. *Nat. Hum. Behav.* **2**, 117–125 (2018).
154. Sanderson, E., Davey Smith, G., Bowden, J. & Munafò, M. R. Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. *Nat. Commun.* **10**, 2949 (2019).
155. Davies, N. M. et al. Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. *eLife* **8**, e43990 (2019).
156. Tillmann, T. et al. Education and coronary heart disease: mendelian randomisation study. *BMJ* **358**, j3542 (2017).
157. Davies, N. M., Dickson, M., Davey Smith, G., Windmeijer, F. & van den Berg, G. J. The causal effects of education on adult health, mortality and income: evidence from Mendelian randomization and the raising of the school leaving age. Preprint at *SSRN* https://doi.org/10.2139/ssrn.3390179 (2019).
158. Baldwin, J., Pingault, J.-B., Schoeler, T., Sallis, H. M. & Munafo, M. R. Protecting against researcher bias in secondary data analysis: challenges and solutions. *Eur. J. Epidemiol.* **37**, 1–10 (2022).
159. Sallis, H. Triangulation protocol; intergenerational effects of parental substance use on child substance use and mental health outcomes. Preprint at https://osf.io/s6jv4/ (2021).
160. Hartwig, F. P., Davies, N. M. & Davey Smith, G. Bias in Mendelian randomization due to assortative mating. *Genet. Epidemiol.* **42**, 608–620 (2018).
161. Brumpton, B. et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat. Commun.* **11**, 3519 (2020).
162. Morris, T. T., Davies, N. M., Hemani, G. & Davey Smith, G. Population phenomena inflate genetic associations of complex social traits. *Sci. Adv.* **6**, eaay0328 (2020).
163. Minică, C. C., Boomsma, D. I., Dolan, C. V., de Geus, E. & Neale, M. C. Empirical comparisons of multiple Mendelian randomization approaches in the presence of assortative mating. *Int. J. Epidemiol.* **49**, 1185–1193 (2020).
164. Davies, N. M. et al. Within family Mendelian randomization studies. *Hum. Mol. Genet.* **28**, R170–R179 (2019).
165. Minică, C. C., Dolan, C. V., Boomsma, D. I., de Geus, E. & Neale, M. C. Extending causality tests with genetic instruments: an integration of Mendelian randomization with the classical twin design. *Behav. Genet.* **48**, 337–349 (2018).
166. Howe, L. J. et al. Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nat. Genet.* (in the press).
167. Taylor, A. E. et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **47**, 1207–1216 (2018).
168. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
169. Pirastu, N. et al. Genetic analyses identify widespread sex-differential participation bias. *Nat. Genet.* **53**, 663–671 (2021).
170. Smit, R. A., Trompet, S., Dekkers, O. M., Jukema, J. W. & le Cessie, S. Survival bias in Mendelian randomization studies: a threat to causal inference. *Epidemiology* **30**, 813 (2019).
171. Schooling, C. M. et al. Use of multivariable Mendelian randomization to address biases due to competing risk before recruitment. *Front. Genet.* **11**, 610852 (2020).
172. Vansteelandt, S., Dukes, O. & Martinussen, T. Survivor bias in Mendelian randomization analysis. *Biostatistics* **19**, 426–443 (2017).
173. Hernán, M. A. Invited commentary: selection bias without colliders. *Am. J. Epidemiol.* **185**, 1048–1050 (2017).
174. Mahmoud, O., Dudbridge, F., Davey Smith, G., Munafo, M. & Tilling, K. Slope-Hunter: a robust method for index-event bias correction in genome-wide association studies of subsequent traits. *Nat. Commun.* (in the press).
175. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).
176. Debat, V. & David, P. Mapping phenotypes: canalization, plasticity and developmental stability. *Trends Ecol. Evol.* **16**, 555–561 (2001).
177. Kitami, T. & Nadeau, J. H. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nat. Genet.* **32**, 191–194 (2002).
178. Gu, Z. et al. Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
179. Hemani, G. et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. Preprint at *bioRxiv* https://doi.org/10.1101/173682 (2017).
180. Ioannidis, J. P. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q.* **94**, 485–514 (2016).
181. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
182. Paternoster, L., Tilling, K. & Davey Smith, G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: conceptual and methodological challenges. *PLoS Genet.* **13**, e1006944 (2017).
183. Zhou, W. et al. Causal relationships between body mass index, smoking and lung cancer: univariable and multivariable Mendelian randomization. *Int. J. Cancer* **148**, 1077–1086 (2021).
184. Lee, J. C. et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat. Genet.* **49**, 262–268 (2017).
185. Kim, Y.-I. Role of folate in colon cancer development and progression. *J. Nutr.* **133**, 3731S–3739S (2003).
186. Davey Smith, G., Paternoster, L. & Relton, C. When will Mendelian randomization become relevant for clinical practice and public health? *JAMA* **317**, 589–591 (2017).
187. Ye, T., Shao, J. & Kang, H. Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization. *Ann. Stat.* **49**, 2079–2100 (2021).
188. Bowden, J. et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *Int. J. Epidemiol.* **48**, 728–742 (2018).
189. Wang, S. & Kang, H. Weak-instrument robust tests in two-sample summary-data Mendelian randomization. *Biometrics* https://doi.org/10.1111/biom.13524 (2021).
190. Minelli, C. et al. The use of two-sample methods for Mendelian randomization analyses on single large datasets. *Int. J. Epidemiol.* **50**, 1651–1659 (2021).
191. Burgess, S., Foley, C. N., Allara, E., Staley, J. R. & Howson, J. M. M. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nat. Commun.* **11**, 376 (2020).
192. Foley, C. N., Mason, A. M., Kirk, P. D. W. & Burgess, S. MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates. *Bioinformatics* **37**, 531–541 (2020).
193. Berzuini, C., Guo, H., Burgess, S. & Bernardinelli, L. A Bayesian approach to Mendelian randomization with multiple pleiotropic variants. *Biostatistics* **21**, 86–101 (2018).
194. Xu, S., Fung, W. K. & Liu, Z. MRCIP: a robust Mendelian randomization method accounting for correlated and idiosyncratic pleiotropy. *Brief. Bioinform*. **22**, bbab019 (2021).
195. Qi, G. & Chatterjee, N. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nat. Commun.* **10**, 1941 (2019).
196. Cheng, Q. et al. MR-LDP: a two-sample Mendelian randomization for GWAS summary statistics accounting for linkage disequilibrium and horizontal pleiotropy. *NAR Genom. Bioinform* **2**, lqaa028 (2020).
197. Zhu, X., Li, X., Xu, R. & Wang, T. An iterative approach to detect pleiotropy and perform Mendelian randomization analysis using GWAS summary statistics. *Bioinformatics* **37**, 1390–1400 (2020).
198. Grant, A. J. & Burgess, S. An efficient and robust approach to Mendelian randomization with measured pleiotropic effects in a high-dimensional setting. *Biostatistics* https://doi.org/10.1093/biostatistics/kxaa045 (2020).
199. Iong, D., Zhao, Q. & Chen, Y. A latent mixture model for heterogeneous causal mechanisms in mendelian randomization. Preprint at https://arxiv.org/abs/2007.06476 (2020).
200. van der Graaf, A. et al. Mendelian randomization while jointly modeling *cis* genetics identifies causal relationships between gene expression and lipids. *Nat. Commun.* **11**, 4930 (2020).
201. Jiang, L., Xu, S., Mancuso, N., Newcombe, P. J. & Conti, D. V. A hierarchical approach using marginal summary statistics for multiple intermediates in a Mendelian randomization or transcriptome analysis. *Am. J. Epidemiol.* **190**, 1148–1158 (2021).
202. DiPrete, T. A., Burik, C. A. P. & Koellinger, P. D. Genetic instrumental variable regression: explaining socioeconomic and health outcomes in nonexperimental data. *Proc. Natl Acad. Sci. USA* **115**, E4970–E4979 (2018).
203. Howey, R., Shin, S.-Y., Relton, C., Davey Smith, G. & Cordell, H. J. Bayesian network analysis incorporating genetic anchors complements conventional Mendelian randomization approaches for exploratory analysis of causal relationships in complex data. *PLoS Genet.* **16**, e1008198 (2020).
204. Schmidt, A. F. & Dudbridge, F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. *Int. J. Epidemiol.* **47**, 1217–1228 (2017).
205. Bucur, I. G., Claassen, T. & Heskes, T. Inferring the direction of a causal link and estimating its

effect via a Bayesian Mendelian randomization approach. *Stat. Methods Med. Res.* **29**, 1081–1111 (2019).
206. Davey Smith, G., Holmes, M. V., Davies, N. M. & Ebrahim, S. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *Eur. J. Epidemiol.* **35**, 99–111 (2020).
207. Davey Smith, G. et al. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med.* **4**, 1985–1992 (2007).
208. Pearl, J. *Causality* (Cambridge Univ. Press, 2009).
209. Keele, L., Zhao, Q., Kelz, R. R. & Small, D. Falsification tests for instrumental variable designs with an application to tendency to operate. *Med. Care* **57**, 167–171 (2019).
210. Brookhart, M. A., Rassen, J. A. & Schneeweiss, S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol. Drug Saf.* **19**, 537–554 (2010).
211. Burgess, S. & Labrecque, J. A. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur. J. Epidemiol.* **33**, 947–952 (2018).
212. Wang, L. & Tchetgen Tchetgen, E. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J. R. Stat. Soc. Ser. B* **80**, 531–550 (2018).
213. Mills, H. L. et al. Detecting heterogeneity of intervention effects using analysis and meta-analysis of differences in variance between arms of a trial. *Epidemiology* **32**, 846–854 (2021).
214. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, 2003).
215. Zuckerkandl, E. & Villet, R. Concentration-affinity equivalence in gene regulation: convergence of genetic and environmental effects. *Proc. Natl Acad. Sci. USA* **85**, 4784–4788 (1988).
216. Ebrahim, S. & Davey Smith, G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum. Genet.* **123**, 15–33 (2008).
217. Hill, W. D. et al. Molecular genetic contributions to social deprivation and household income in UK Biobank. *Curr. Biol.* **26**, 3083–3089 (2016).
218. Labrecque, J. A. & Swanson, S. A. Interpretation and potential biases of mendelian randomization estimates with time-varying exposures. *Am. J. Epidemiol.* **188**, 231–238 (2018).
219. Sanderson, E., Richardson, T. G., Morris, T. T., Tilling, K. & Davey Smith, G. Estimation of causal effects of a time-varying exposure at multiple time points through Multivariable Mendelian randomization. Preprint at *medRxiv* https://doi.org/10.1101/2022.01.04.22268740 (2022).
220. Cardon, L. R. & Palmer, L. J. Population stratification and spurious allelic association. *Lancet* **361**, 598–604 (2003).
221. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284 (2015).
222. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
223. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
224. Lawson, D. J. et al. Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Hum. Genet.* **139**, 23–41 (2020).
225. Haworth, S. et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.* **10**, 1–9 (2019).
226. Howe, L. J. et al. Genetic evidence for assortative mating on alcohol consumption in the UK Biobank. *Nat. Commun.* **10**, 5039 (2019).
227. Nordsletten, A. E. et al. Patterns of nonrandom mating within and across 11 major psychiatric disorders. *JAMA Psychiat.* **73**, 354–361 (2016).
228. Bochud, M., Chiolero, A., Elston, R. C. & Paccaud, F. A cautionary note on the use of Mendelian randomization to infer causation in observational epidemiology. *Int. J. Epidemiol.* **37**, 414–416 (2008).

Acknowledgements

E.S., M.R.M., T.P. and G.D.S. are members of the UK Medical Research Council (MRC) Integrative Epidemiology unit, which is funded by the MRC (MC_UU_00011/1, MC_UU_00011/3 and MC_UU_00011/7) and the University of Bristol. M.M.G. is supported by the National Institutes of Health/National Institute on Aging (NIH/NIA) grant R01AG057869. M.V.H. works in a unit that receives funding from the MRC and is supported by a British Heart Foundation Intermediate Clinical Research Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. H.K. is supported by the National Science Foundation grant DMS-1811414. C.W. is funded by the MRC (MC UU 00002/4, MC UU 00002/13) and the Wellcome Trust (WT107881).

Author contributions

Introduction (E.S.); Experimentation (E.S., M.M.G. and T.P.); Results (E.S., M.M.G., T.P. and C.W); Applications (E.S. and M.V.H.); Reproducibility and data deposition (M.R.M.); Limitations and optimizations (E.S.); Outlook (G.D.S.); Overview of the Primer (E.S., H.K., J.M., C.M.S., Q.Z. and G.D.S.).

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Methods Primers thanks Marianne Benn, Frida Emanuelsson, Sarah Gagliano Taliun and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

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

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1038/s43586-021-00092-5.

RELATED LINKS

ivonesamplemr: https://github.com/remlapmot/ivonesamplemr
MendelianRandomization: https://cran.r-project.org/package=MendelianRandomization
MR dictionary: https://mr-dictionary.mrcieu.ac.uk/
mrrobust: https://github.com/remlapmot/mrrobust
OneSampleMR: https://remlapmot.github.io/OneSampleMR/
STROBE-MR: https://www.strobe-mr.org/
The OpenGWAS project: https://gwas.mrcieu.ac.uk/
TwoSampleMR: https://github.com/MRCIEU/TwoSampleMR
UK Biobank: https://www.ukbiobank.ac.uk/