VU Research Portal

Understanding the assumptions underlying Mendelian randomization

de Leeuw, Christiaan; Savage, Jeanne; Bucur, Ioan Gabriel; Heskes, Tom; Posthuma, Danielle

Link to publication in VU Research Portal

REVIEW ARTICLE

Understanding the assumptions underlying Mendelian randomization

Christiaan de Leeuw [1] ✉, Jeanne Savage [1], Ioan Gabriel Bucur [2], Tom Heskes[2] and Danielle Posthuma [1,3]

With the rapidly increasing availability of large genetic data sets in recent years, Mendelian Randomization (MR) has quickly gained popularity as a novel secondary analysis method. Leveraging genetic variants as instrumental variables, MR can be used to estimate the causal effects of one phenotype on another even when experimental research is not feasible, and therefore has the potential to be highly informative. It is dependent on strong assumptions however, often producing biased results if these are not met. It is therefore imperative that these assumptions are well-understood by researchers aiming to use MR, in order to evaluate their validity in the context of their analyses and data. The aim of this perspective is therefore to further elucidate these assumptions and the role they play in MR, as well as how different kinds of data can be used to further support them.

INTRODUCTION

Genetic research in the last two decades has taken an enormous flight, and a wealth of genetic data is now available for a wide variety of human phenotypes [1]. Besides providing ever-increasing insight into the genetic etiology of these phenotypes, it may provide an opportunity to study causal relations between these phenotypes as well.

Although causal inference is generally considered the domain of experimental methods like randomized controlled trials (RCT), some nonexperimental methods can be applied to estimate causal relations indirectly [2]. Though less robust, these can be used when RCTs are not a viable option. Mendelian Randomization (MR), a form of instrumental variable analysis that uses genetic variants as instruments to investigate causal relations between phenotypes, is one such method [3]. MR has become very popular in recent years, with thousands of methodological and applied MR studies published to date [4, 5], and with the continued growth of available genetic data this trend will likely persist.

MR relies on strong assumptions however, yielding biased and misleading results if those assumptions fail [6, 7]. Given the widespread popularity of MR, it is therefore imperative that these assumptions are clearly understood by the researchers using it, to allow them to properly evaluate the validity of these assumptions in the context of their own data and analyses [8–10].

The aim of this Perspective is to outline the assumptions that are needed to perform MR, what role those assumptions play in the analysis and its interpretation, and what information different elements of input data contribute to the support of these assumptions. Our aim is not to give an exhaustive overview of individual methods, but rather to elucidate the underlying logic of MR in its different forms. As such, we will also abstract away from issues pertaining to estimation, assuming an idealized scenario in which all associations between observed variables are fully known, examining what challenges remain even when estimation uncertainty is entirely eliminated.

CORE PRINCIPLE

The aim of an MR analysis is to estimate and test the causal effect of a putative causal phenotype X, the exposure, on another phenotype Y, the outcome. It uses the principles of instrumental variable analysis to do so, with the genotype G_j of a genetic variant j serving as the instrument [8, 11].

To serve as a valid instrument for the causal effect of exposure on outcome, there must be an association between G_j and the exposure. Moreover, it must be the case that any association of G_j with the outcome is mediated by the exposure, as depicted in Fig. 1A. In other words, associations of G_j directly with the outcome, or with a variable C that acts as a confounder of exposure and outcome cannot be present (Fig. 1B). There is no requirement that G_j itself has a causal effect (see also *Supplementary Information—Relevance assumption*); if variant j is in LD with causal variants that are valid instruments, then G_j is a valid instrumental variable as well (Fig. 1C). For ease of notation however, the graphs used throughout the paper will assume the selected variants used are causal.

If we assume the effect sizes of all associations and causal effects to be constant (i.e., simple linear relations), we can easily see how this can provide the parameter β_{XY} of the causal effect of the exposure on the outcome. Denoting the marginal associations of G_j with exposure and outcome as γ_{Xj} and γ_{Yj} respectively, for the assumed scenario in Fig. 1A we can express these as $\gamma_{Xj} = \alpha_{Xj}$ and $\gamma_{Yj} = \alpha_{Xj}\beta_{XY}$. Because the association γ_{Yj} between G_j and the

[1]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, The Netherlands. [2]Department of Data Science, Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands. [3]Department of Clinical Genetics, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, The Netherlands. ✉email: c.a.de.leeuw@vu.nl
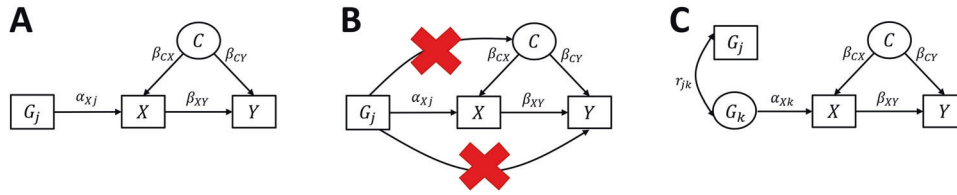
Fig. 1 Graphical representation of valid instrument causal scenarios, for a variant *j*. These causal graphs depict the genetic instrumental variable assumptions on which MR is based, with a genetic variant with genotype G_j causally affecting the exposure X which in turn (potentially) causally affects Y, while allowing for the presence of confounders C of the exposure and outcome. In this and subsequent figures, variables are shown as rectangles or ovals, with ovals denoting that the variable is not (necessarily) observed, and causal effects are indicated using one-sided arrows in the direction of the causal effect, with an accompanying effect size parameter shown next to it. Two-sided arrows denote correlations between variables caused by other variables external to the model. For simplicity of notation throughout the paper, all variables are assumed to be standardized, with mean of zero and unit variance. Shown in (**A**) is the basic valid instrument scenario, with in (**B**) the same graph emphasizing the causal paths explicitly ruled out by the independence and exclusion restriction assumptions. The graph in (**C**) shows an alternative valid instrument scenario where the variant *j* used is not causal, but is in LD with another variant *k* that is.

Table 1. Instrumental variable and other assumptions relevant for MR.

Assumption	Description
Instrumental variable assumptions	
Relevance	The variant is associated with the outcome ($\gamma_{Xj} \neq 0$); the variant does not need to be causal
Independence	The variant is not associated with any confounders ($a_{Cj} = 0$)
Exclusion restriction	The variant is independent of the outcome given the exposure and all confounders ($a_{Yj} = 0$)
Additional assumptions	
Constant effect sizes	
Same population parameters (multi-sample)	For multi-sample analyses, the (relevant) parameters are the same across all populations the different cohorts were drawn from
Same conditioning	The associations used are all conditioned on (relevantly) the same variables and in the same way, in terms of covariates included in analyses as well as selection effects (in multi-sample analysis)
No nonlinearities	Effect sizes for any causal effect or association are not dependent on the value of either of the two variables (as opposed to e.g., quadratic effect of causal variable, or with a binary outcome)
No interaction effects	Effect sizes for any causal effect or association are not dependent on the value of any other variable
Fully observed variables	The observed instance of each variable fully reflects the causally relevant instance of that variable; that is, it is observed without noise or rescaling relative to the causal instance

Which assumptions are required for a given MR analysis depends on the model used (see text).

outcome is fully mediated by the exposure, it equals the causal effect β_{XY} scaled by the causal effect a_{Xj} of G_j on the exposure.

Thus, defining the ratio of marginal effects $\beta_j = \frac{\gamma_{Yj}}{\gamma_{Xj}}$, it follows that if variant *j* is a valid instrument then $\beta_j = \frac{a_{Xj}\beta_{XY}}{a_{Xj}} = \beta_{XY}$ [11]. In other words, the variant-specific causal effect a_{Xj} cancels out in the ratio of the marginal genetic effects, making β_j equal to the causal effect parameter β_{XY} for every variant that is a valid instrument. Although not every MR method is explicitly defined in terms of β_j, they all ultimately depend on this property. To examine the impact of different causal scenarios, we will thus focus on the functional form β_j takes in those scenarios, and whether it still equals β_{XY}.

We can thus obtain β_{XY} using any genetic variant for which the instrumental variable assumptions hold [12], since all such variants provide the same causal parameter. However, the a priori plausibility of these assumptions varies greatly, depending particularly on the exposure being studied, and establishing that the variants used are indeed valid instruments requires further analysis and data. As such it is crucial that active steps are taken to ensure that all assumptions are met, since reliable interpretation of MR results is otherwise impossible.

MR also generally depends on some additional assumptions [8, 13], which are listed in Table 1. Different methods may relax these additional assumptions in various ways so these are not always all required. In the next two sections, we will examine causal scenarios that violate the instrumental variable assumptions, and various

strategies to deal with such violations, either by direct modeling and testing or by levering constrained data. Following that we discuss the role of the additional assumptions and what can happen if they do not hold. Throughout, we will use the simplest causal scenario that can illustrate the particular issue being discussed, rather than providing an exhaustive list of such scenarios. Additional discussion and mathematical details for these issues is found in the Supplemental Information. An overview of the main methods referenced is given in Table 2.

EVALUATING INSTRUMENTAL VARIABLE ASSUMPTIONS
Heterogeneity of causal estimates
One common way in which the exclusion restriction can be violated is by a direct causal effect of the genetic variant on the outcome (Fig. 2A). The reason why this is a problem can be readily discerned when considering how this changes the functional form of the marginal association γ_{Yj} of the variant with the outcome, which becomes $\gamma_{Yj} = a_{Xj}\beta_{XY} + a_{Yj}$ This means that the ratio parameter β_j now equals $\beta_j = \frac{a_{Xj}\beta_{XY} + a_{Yj}}{a_{Xj}} = \beta_{XY} + \frac{a_{Yj}}{a_{Xj}}$. The same thing happens in a scenario where there is LD between G_j and another variant G_k that has a causal effect on the outcome (Fig. 2B).

In other words, β_j becomes offset from the value of the true causal effect β_{XY} by a bias term specific to that variant. Although in this case we can no longer directly obtain the causal effect from β_j, the way this type of violation manifests itself makes it relatively

Table 2. Overview of referenced methods.

Method	Brief description
Basic multivariant MR methods	
Two-stage least squares [3]	General instrumental variable analysis model for single-sample MR
IVW mean [3]	Estimates inverse-variance weighted (IVW) mean of the β_j
Heterogeneity testing	
GSMR [14]	Combination of IVW mean with HEIDI heterogeneity test
GLIDE [15]	Heterogeneity test, using set of simultaneous regression equations
MR-PRESSO [16]	Heterogeneity test, using discrepancy between each variant and IVW estimate based on rest of variants
Implicit subset MR methods	
Bowden et al. [17]	Estimates weighted median of the β_j
Hartwig et al. [18]	Estimates weighted mode of the β_j using empirically smoothed densities
Burgess et al. [19]	Estimates weighted mode of the β_j using heterogeneity weighted average density of IVW estimates of all subsets of variants
MR-Mix [20]	Models the set variants as an implicit mixture of valid and invalid instruments, and derives the estimate from the valid component of the mixture
Modeled pleiotropy MR methods	
MR-Egger [21]	Estimation via weighted linear regression of γ_{Yj} on γ_{Xj}
BayesMR [22]	Bayesian model selection on forward and reverse causation models
CAUSE [24]	Bayesian mixture model allowing a subset of variants to correspond to a mediated confounding scenario (whole-genome analysis)
LHC-MR [23]	Mixture model allowing different subsets of variants to correspond to mediated confounding and reverse causation scenarios (whole-genome analysis)
Explicit confounder MR methods	
Multivariable MR-Egger [26]	MR-Egger approach that includes additional γ_{Cj} in the model
MR-TRYX [25]	Large-scale evaluation of potential confounding using GWAS summary statistics database
Negative control population MR methods	
PRMR [32]	Estimates the total component of γ_{Yj} not mediated by X using a negative control population

straightforward to detect. Because this bias term is variant-specific it will tend to differ across (independent) variants, resulting in a heterogeneity of their β_j values (see also *Supplementary Information—heterogeneity of estimated causal effects*). By contrast, for a set of variants that are all valid instruments, their β_j will be the same, because as noted above they will all equal the causal effect parameter β_{XY}.

Given this, if we have multiple variants available as potential genetic instruments, an obvious and commonly used way to leverage this is therefore to test for heterogeneity of the β_j. Then, if such heterogeneity is found to be present, we can prune away variants from the selection until we retain a subset of variants with homogeneous β_j. In this way we can rule out violations of the exclusion restriction of the kind depicted in Fig. 2A, B, and under the assumption that the remaining variants are valid instruments we can use those variants to obtain β_{XY} as before [14–16].

An alternative to explicit heterogeneity testing and pruning is to use "robust" models for multivariant MR analysis, which do not require that all variants used for their input are valid instruments (see also *Supplementary Information—robust methods*). These subdivide into two main types. The first type assumes that only a subset of the variants used are valid instruments, and take either a median- or mode-based approach. Median-based methods only require that more than half of the variants are valid instruments, which guarantees that the median of the β_j equals β_{XY} [17]. Mode-based methods make an even weaker assumption, only requiring that the largest subset of variants with homogeneous β_j consists of valid instruments, in which case the mode of the β_j will equal β_{XY} [18–20].

The second type of robust model does not require that any variant is a valid instrument. Instead, it models the marginal

association of each variant with the outcome as $\gamma_{Yj} = \gamma_{Xj}\beta_{XY} + \delta_j$ with a heterogeneity term δ_j, and then makes an assumption about the distribution of these δ_j. The most prominent example of this second type is the MR-Egger model [21], which is based on the so-called InSIDE (Instrument Strength Independent of Direct Effect) assumption. This assumption states that these δ_j terms are independent of the marginal associations γ_{Xj} of the variant with the exposure, and based on this the MR-Egger model can estimate β_{XY} using essentially a linear regression of γ_{Yj} on γ_{Xj}. For valid instruments this assumption is automatically true, since δ_j is zero, and for a scenario such as in Fig. 2A it is very plausible as well: in that case, $\gamma_{Xj} = \alpha_{Xj}$ and $\delta_j = \alpha_{Yj}$, and since α_{Xj} and α_{Yj} represent two distinct causal paths that share no mediating variables there is no clear mechanism by which they would become correlated.

Robust methods can thus in principle directly estimate the causal effect from a mixture of valid and invalid instruments, but this requires specific assumptions about the degree or structure of the heterogeneity, which are not directly testable. Even when using such robust methods, it is therefore still imperative that the heterogeneity, and the validity of the assumptions made about it (with specific valid subsets of variants present in the data for median- and mode-based methods, or the independence specified by InSIDE for MR-Egger), are explicitly considered.

Moreover, homogeneity of the β_j does not imply that the instrumental variable assumptions (or the InSIDE assumption) do hold, since there are other causal scenarios that violate the assumptions without resulting in heterogeneity. For the remainder of the paper, we will therefore generally assume that heterogeneity has been dealt with, and focus on scenarios where all variants used correspond to the same homogeneous causal graph, and with β_j equal to the same value β.
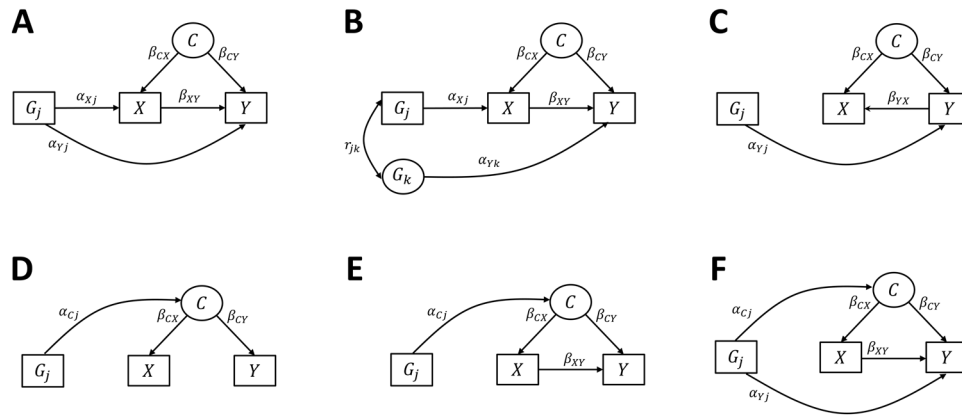
Fig. 2 Graphical representation of several violations of instrumental variable assumptions, for a variant *j*. In (**A**, **B**) are two similar violations of the exclusion restriction, with causal effects directly on the outcome either from variant *j* itself or from another variant *k* in LD with it. **C** Shows a reverse causation scenario, another violation of the exclusion restriction, with a causal effect of variant *j* directly on the outcome, which is then mediated onto the exposure by the causal effect of outcome on exposure. **D**, **E** show two mediated confounding scenarios which violate the independence assumption, with the confounder *C* mediating the genetic effect of variant *j* onto both the exposure and outcome, with in (**F**) a further variation on (**E**) with additional direct causal effects of the variant on the outcome.

Reverse causation

The "reverse causation" scenario is illustrated in Fig. 2C, the mirror image of Fig. 1A, with the genetic variant now exerting a direct causal effect on the outcome, which in turn has a causal effect on the exposure. This is also a violation of the exclusion restriction, but unlike in Fig. 2A, B this does not result in heterogeneity. This is because the marginal genetic associations of the variant are $\gamma_{Xj} = \alpha_{Yj}\beta_{YX}$ and $\gamma_{Yj} = \alpha_{Yj}$, which means that $\beta = \frac{\alpha_{Yj}}{\alpha_{Yj}\beta_{YX}} = \frac{1}{\beta_{YX}}$, the inverse of the causal effect of the outcome on the exposure. As such, the value of β we would get in this scenario is completely different from the β_{XY} we are attempting to estimate, which in this case is simply zero. The InSIDE assumption also does not hold here, since the heterogeneity term $\delta_j = \alpha_{Yj}$, meaning that both δ_j and γ_{Xj} are dependent on the same parameter α_{Yj}.

When the genetic effect on the outcome is fully mediated by the exposure as in Fig. 1A, it follows that the correlations between the variant and the outcome are weaker than those between the variant and the outcome; unless the exposure fully determines the outcome in which case the correlations are equal. In case of reverse causation, as in Fig. 2C, the opposite is true, with the correlations between variant and exposure being weaker than those between variant and outcome. For Fig. 1A, since in our notation all variables are standardized, the correlations of the variant with the exposure and outcome equal the genetic associations γ_{Xj} and γ_{Yj} respectively, and the standardization also means that the absolute value of all causal parameters is at most one as well, including β_{XY}. Since as previously noted $\gamma_{Yj} = \gamma_{Xj}\beta_{XY}$, the absolute value of γ_{Yj} must therefore be smaller than (or at most equal to) that of γ_{Xj}.

It is therefore generally possible to infer direction from the relative size of these correlations, or more directly from the causal estimate itself. In case of reverse causation $\beta_j = \frac{1}{\beta_{YX}}$, which (since $|\beta_{YX}|$ is at most 1) will have an absolute value greater than or equal to 1. As such we can decide between forward and reverse causation by determining whether β_j is smaller or greater than 1. This can be assessed manually by running MR analyses in both directions or using a model that incorporates both [22, 23]. Moreover, depending on the choice of exposure and outcome we will often already have strong a priori information about the causal direction, and in some cases reverse causation is inherently impossible because the exposure is known to occur before the outcome. In this regard, resolving the order of causation is often relatively straightforward in practice.

However, these methods and a priori information can only help to decide between forward and reverse causation as long as the

independence assumption holds, and it is thus presumed that one of these two scenarios is correct. This therefore still requires ruling out the possibility of genetic effects on exposure and outcome being mediated by one of their confounders.

Analysing potential confounders

Two variations of what we will refer to as "mediated confounding" are depicted in Fig. 2D, E, with a causal effect α_{Cj} of the variant on a confounder *C*, violating the independence assumption. These scenarios result in a β value of $\beta_{XY} + \frac{\beta_{CY}}{\beta_{CX}}$ (with $\beta_{XY} = 0$ for Fig. 2D), demonstrating a bias away from the true causal effect of the exposure on the outcome. The InSIDE assumption is violated here as well, with both $\gamma_{Xj} = \alpha_{Cj}\beta_{CX}$ and $\delta_j = \alpha_{Cj}\beta_{CY}$ dependent on α_{Cj}. Note that these scenarios are specific to the particular confounder *C*, and there may be other sets of variants operating on different confounder variables, with correspondingly different biases.

Because the $\beta_{XY} + \frac{\beta_{CY}}{\beta_{CX}}$ term can take any value that β_{XY} itself can take, it is impossible to rule out mediated confounding scenarios using just the genetic associations with exposure and outcome. Some methods have been developed that use a mixture model approach to explicitly include a mediated confounding component in their model, such as CAUSE [24] which assumes that the variants used are a mixture of ones conforming to Fig. 2A and others conforming to Fig. 2F. LHC-MR [23] offers an even more general model also allowing for reverse causation. However, the problem remains that for any forward causation scenario as in Fig. 2A, it is possible to formulate parameter values for the mediated confounding scenario like Fig. 2F that result in an identical pattern of genetic associations. As such, the components of these mixture models that are assumed to capture forward causation may still be capturing mediated confounding instead (see also *Supplementary Information—whole-genome methods*).

Additional data is therefore required to resolve the issue of mediated confounding. If genetic associations conditioning on a putative confounder variable *C* are available for both exposure and outcome, evaluating and correcting for that particular *C* is relatively straightforward. If this *C* is indeed mediating (part of) the effect of the variants on the exposure and outcome, adding *C* as a covariate to compute the conditional associations will remove this confounding effect from a subsequent MR analysis based on them. Similarly, if separate GWAS results for a possible confounder *C* are available, these can be used to obtain corrected MR estimates. This can be accomplished by either first correcting the γ_{Xj} and γ_{Yj} and then performing a regular MR analyis [25], or by using an MR-Egger style regression approach, essentially
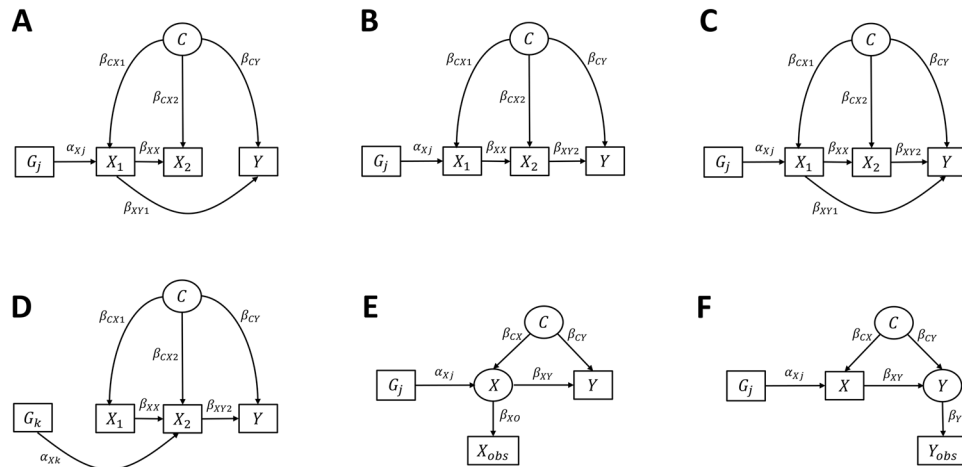
Fig. 3 Graphical representation of scenarios involving longitudinal data and imperfect measurement of variables, for a variant *j*.
Different longitudinal data scenarios are shown in (**A**) through (**D**), with X_1 and X_2 corresponding to an earlier and later measurement of the exposure. Causal effects on the outcome occur either (**A**) at the earlier time point, (**B**) the later time point, or (**C**) at both time points, with in (**D**) an additional scenario where variant *j* directly affects only the later measurement of the exposure and not the earlier one. In (**E**) is shown a scenario where the observed exposure X_{obs} does not fully represent the causally relevant instance X of the exposure, and the same in (**F**) for an observed outcome Y_{obs} that does not fully represent the causally relevant outcome Y.

regressing γ_{Yj} on both γ_{Xj} and γ_{Cj} (the genetic associations with the possible confounder) simultaneously. The latter approach can be considered a form of multiple-exposure model, treating *C* as a second exposure potentially correlated with *X* [26]. Note that both correction using *C* directly or based on the γ_{Cj} is susceptible to collider bias when *C* is not a confounder [27], which therefore needs to be considered when using such methods (see also *Supplementary Information—mediated confounding*).

Although approaches like these can be effective in detecting and correcting for effects mediated by confounders, the obvious limiting factor is that this requires the potential confounders to be explicitly tested. If no data is available for a particular confounder, or if it was simply not considered as a potential confounder in the analysis, its effects will not have been accounted for. This poses a major challenge, since any confounder of the exposure and outcome is itself almost certainly heritable, and any variant directly associated with that confounder will also have associations with the exposure and outcome mediated by that confounder.

This implies that in practice all (potential) confounders of the exposure and outcome would need to be considered and evaluated in an MR context. This is particularly problematic with confounding endophenotypes such as those involved in specific biological pathways and processes, as their causal effects on exposure and outcome may be specific to a particular context such as a tissue or developmental time period, and measurements of such confounders would therefore need to be specific to that context as well.

LEVERAGING CONSTRAINED DATA
Negative control populations
MR has sometimes been compared to RCTs, drawing a parallel between the random inheritance of alleles from parents to offspring and the randomized assignment of study participants to treatment groups, with the exposure taking the role that the actual treatment has in RCT [28]. However, this analogy is problematic, because although part of the inferential strength of RCT comes from random assignment of individuals to groups, such randomization only deals with pre-existing differences between individuals in the trial. Potential confounding that occurs after assignment remains a constant challenged even in RCT and must accounted for in the experimental design, by using well-designed control groups and strictly controlling other

experimental and background variables. This level of control does not exist in the MR context, and since the exposure occurs at an unknown time possibly many years after the "randomized assignment" (and measurement of the exposure and outcome typically happens even later still), there is ample opportunity for confounding to arise.

An MR approach that more closely mimics the structure of RCT however, is the use of negative control populations [13, 29]. A negative control population is one where the exposure is constrained to a particular value, but that in other respects matches the population from which the main MR data for was derived (i.e., the relations between all relevant variables are the same). An example of this is alcohol consumption as the exposure, using a population where people do not drink alcohol due to religious or cultural taboo as control [30]. A negative control population does need to have an actual constraint on the exposure; simply selecting a subset of a population for whom the exposure is zero does not work, as this would lead to collider bias (see *Supplementary Information—negative control populations*).

Because in such a control population the exposure does not vary, causal effects involving that exposure are essentially blocked. The constraint on the exposure stops other variables from affecting the exposure, and stops the exposure from affecting other variables. Genetic association between a variant and the outcome in this control population therefore only consists of effects not mediated by the exposure, and thus should be zero for valid genetic instruments like in Fig. 1A. Testing the genetic association between variants and the outcome can thus serve to validate them as instruments, provided the control sample is sufficiently well-powered.

This approach can be further extended to determine how much of the genetic association with the outcome γ_{Yj} is not mediated by the exposure (with some restrictions, see *Supplementary Information—negative control populations*) [31]. Modeling this genetic association as $\gamma_{Yj} = \gamma_{Xj}\beta_{XY} + \delta_j$, similar to MR-Egger, this can essentially provide a direct estimate of the heterogeneity term δ_j for each individual variant *j*. With that, it becomes possible to obtain a corrected genetic association $\gamma_{Yj} - \delta_j$, by subtracting out the heterogeneity from the overall association, and then using this corrected γ_{Yj} to perform MR analysis. However, although potentially quite powerful, using negative control populations in this way is also vulnerable to bias, since this will create a hidden bias if the assumptions of the negative control population fail. This is in

contrast to using negative control populations to determine validity of variants as an instrument, which will instead only tend to generate false negatives (rejecting valid instruments as invalid) if the negative control population assumptions do not hold.

Other forms of constrained data

Using negative control populations leverages natural constraints on data to provide a means of validating the instrumental variable assumptions that does not require explicit testing of individual confounders. Other approaches that utilize such constraints can be employed as well, and a prime example of this is the use of longitudinal data, for either exposure, outcome, or both. Use of such data allows the timing of the causally relevant exposure and of the causal effects to be narrowed down.

If for example we have two measurements of the exposure, as in Richardson et al. [32], there are three main scenarios to consider: a direct causal effect on the outcome only by the early exposure X_1 (Fig. 3A), only by the late exposure X_2 (Fig. 3B), or by both (Fig. 3C). This can be resolved by a set of three MR analyses, including one that has X_2 as the exposure with a set of variants such as in Fig. 3D that only affect the later exposure. Here, the early exposure essentially functions as a baseline value, allowing us to identify variants that only affect the change in exposure that occurred since the first time point (see also *Supplementary Information—longitudinal data*).

This process can be generalized to more than two time points, allowing for better determination of the likely timing of the causal effects. If longitudinal measurements of the outcome are available, these can be used in the same way to narrow down the timing. Moreover, for later time points these models can be interpreted as conditioning on the value of the exposure or outcome at an earlier time point, which would block any confounder-mediated genetic effects that occurred prior to that time point from affecting the estimate of β_{XY2} [33]. Although confounders may still be present for the later time points (acting e.g., on X_2 and Y in Fig. 3A), this is restricted to a more limited time window, making it easier to identify likely confounders and correct for them.

Another way of leveraging known constraints on data is the use of positive and negative control outcomes: outcomes which already have strong evidence that they respectively are or are not causally influenced by the exposure, which can be used to evaluate the validity of candidate genetic instruments [8, 34]. Positive control outcomes are subject to a causal effect of the exposure, and as such any variants causally acting on the exposure must be affecting such control outcomes as well. As such, if the variants used in our MR analysis show no association with this positive control outcome, beyond what could be explained by possible lack of statistical power, this suggests that the variants used do not in fact have such a causal effect on the exposure. Similarly, if we perform an MR analysis with a negative control outcome that should not be causally affected by the exposure, and the analysis suggests that there actually is a causal effect on that negative control outcome, this casts doubt on the validity of the variants used as genetic instruments.

RELAXING THE ADDITIONAL ASSUMPTIONS

The causal graph in Fig. 1A is a common way of depicting the instrumental variable assumptions central to MR, clearly showing the causal paths that need to be either present or absent for the standard analysis to work. Less explicit in this graph are some of the additional assumptions implied by it, listed in Table 1, that the analysis depends on as well. These assumptions can be condensed to two general constraints: first, that the causal graph applies in the same way to every individual used in the analysis, both in its structure and in the value of the causal effect sizes; and second,

that the variables as we have measured them in our data, correspond to the true causal variables depicted in the graph without bias or error. In this section we will discuss scenarios in which these assumptions may not hold, and the implications of this for the MR analysis.

Variable effect sizes across samples

In the commonly used two-sample approach to MR analysis, variable effect sizes can potentially occur and pose a problem when the genetic associations γ_{Xj} and γ_{Yj} are obtained from samples each derived from different populations with different values for the causal parameters in Fig. 1A. As described, MR works on the core premise that $\gamma_{Xj} = \alpha_{Xj}$ and $\gamma_{Yj} = \alpha_{Xj}\beta_{XY}$, and that therefore the variant-specific part α_{Xj} will cancel out when we take their ratio $\beta_j = \frac{\gamma_{Yj}}{\gamma_{Xj}}$, leaving only β_{XY}. But this will fail if the value of α_{Xj} in the population from which the exposure GWAS was drawn, differs from the value of α_{Xj} in the population that the outcome GWAS was based on, resulting in β_j being biased away from β_{XY}.

The extent to which this is a problem will depend on the way the MR analysis is conducted. The biases produced by this scenario will usually cause heterogeneity of the β_j, and as such it should be possible to detect and remove the affected variants (see also *Supplementary Information—variable effect sizes*). The MR-Egger style models are more susceptible to this issue, as the average bias will tend to end up in their estimate of β_{XY}, which may go unnoticed unless these are used in conjunction with other types of models. Differences in α_{Cj} across the populations from which GWAS data was drawn will pose similar problems when using additional GWAS data with a putative confounder C as outcome to correct for confounding.

A similar issue can arise even when all data is taken from the same population, if the GWAS samples are subject to explicit or implicit selection criteria. If these criteria differ between the exposure and outcome GWAS, this can lead to the same kind of issue as between different populations described above, if the α_{Xj} differ between the selected subpopulations. Moreover, selection effects occurring in the GWAS sample for the outcome also have the potential to result in collider bias, because selection implicitly conditions on the variables being selected on [27, 35]. For example, the outcome may be measured specifically in older individuals, thus selecting for individuals who have survived to that age [36] and resulting in collider bias if the exposure causally affects life expectancy and there are any confounders of the relation between the exposure and outcome [37] (see also *Supplementary Information—variable effect sizes*). This sort of bias will not generally result in any heterogeneity in the β_j, as it will affect every variant in proportionally the same way. Addressing it will therefore often require identifying relevant selection processes and evaluating whether the specific variables involved may be causing collider bias.

Variable effect sizes within samples

Effect sizes may also vary across individuals within a population, due to for example interactions of causal variants with other variables. In this case, different individuals in the population have a different value of α_{Xj}, depending on their score on the interactor variable. In practice, the genetic associations γ_{Xj} would reflect an average of these different α_{Xj} values across the levels of the interactor variable. The γ_{Yj} are based on this average α_{Xj}, and thus as long as the distribution of the interactor variable is the same in both samples this will still cancel out in the ratio $\beta_j = \frac{\gamma_{Yj}}{\gamma_{Xj}}$. On the other hand, if for example the mean of the interactor is greater in one of the samples, this no longer holds. In that case however, as with the differences in α_{Xj} across samples described above, it should result in heterogeneous β_j, and can therefore be addressed by careful application of heterogeneity testing and modeling.

It is possible for the β_{XY} parameter itself to vary across individuals as well, with different causal effect sizes for different

individuals in the population. This can arise as an interaction effect with another variable but also as a non-linear effect of the exposure, which can be seen as essentially an interaction of the exposure with itself. In effect, the value of β_{XY} that MR would estimate in this case is an average of the different β_{XY} values across the levels of the interactor variable. In this sense, this therefore does not substantially affect the MR analysis, since such an average causal effect is still generally interpretable and informative of the relation between exposure and outcome. It can make it somewhat more difficult to generalize however, since this average β_{XY} would be potentially quite different in other populations if the distribution of the interaction variable in that population substantially differs from that in the population from which the outcome GWAS sample was drawn.

Imperfectly observed variables

In the graphs in Figs. 1 and 2 it is implicitly assumed that the observed variables we use in the GWAS, the exposure and outcome, as well as putative confounder variables we may be trying to evaluate, are sufficiently good proxies for the causally relevant variables. Yet this can fail to be the case for a variety of reasons [38, 39]. There could be simple measurement or diagnostic error, where the observed variables in the data are a noisy representation of the variables of interest. The causal graph in Fig. 3E depicts a scenario like this, with the true exposure of interest X now unobserved, and with a noisy observed exposure variable X_{obs} from which the genetic associations γ_{Xj} are estimated. Such situations often also arise when using binary variables, such as a medical diagnosis or a dichotomized continuous variable (e.g., hypertension as dichotomized blood pressure) [40], where the relevant causal effects are likely related to the underlying biological state rather than with the diagnosis or dichotomized value.

This is can arise from more systematic causes as well. It is possible that the context in which the variable was observed does not sufficiently match that of its causally relevant instance: if for instance we use gene expression as our exposure, it may well be that the tissue in which that gene's expression causally affects the outcome is different from the tissue in which the exposure variable we are using in our analysis is measured. Similarly, there may be differences in timing and developmental period, or environmental triggers, or the observed variable may have a complex internal structure, with the causal effect only pertaining to a subtype or subscale of that variable. In case of large differences between the developmental timing of the causal effect of the exposure and when the exposure was measured, processes such as canalization and behavioral adaptive responses may also have amplified or dampened the changes induced by earlier causal effects [10, 41].

Regardless of the underlying mechanism, in a scenario such as in Fig. 3E where the "true" exposure X is imperfectly represented by the observed exposure X_{obs}, the causal effect we would estimate becomes biased away from β_{XY}. For the exposure the genetic effect changes to $\gamma_{Xj} = \alpha_{Xj}\beta_{XO}$, and as such the ratio $\beta_j = \frac{\gamma_{Yj}}{\gamma_{Xj}}$ becomes $\frac{\beta_{XY}}{\beta_{XO}}$. Depending on the nature of the relation between the "true" and observed variables, the value we get may therefore differ considerably from the true value of β_{XY} (see also *Supplementary Information—imperfectly observed variables*). Note that this issue of imperfectly observed variables is not unique to MR, and would pose a problem even in the context of RCT.

All these same mechanisms can operate on the outcome as well, as depicted in Fig. 3F, in which case β_j will be $\beta_{XY}\beta_{YO}$. Although this does affect interpretation, the value we are estimating does still represent a legitimate causal effect, in contrast to Fig. 3E where the causal structure would be misspecified. If for example our intended outcome is true schizophrenia status, and the Y_{obs} we use is diagnosis of schizophrenia, the causal effect we would obtain is that of our exposure on schizophrenia diagnosis, and as such does have a

meaningful interpretation, even if it does not give us an estimate of the causal effect on true schizophrenia status. In this regard, full observation of the exposure is considerably more crucial than full observation of the outcome.

It should also be noted that a further consequence of such issues is that it may no longer be possible to distinguish forward and reverse causation in the way described above [39], since the parameter constraints upon which this would be based would no longer apply in the same way. Similarly, imperfect observation of a putative confounder C will also tend to render corrections of confounding effects only partially effective, not fully removing the confounding effect. Other approaches for evaluating these alternative causal scenarios would therefore need to be employed.

A somewhat related issue is that even if the observed exposure is in fact a good proxy for the causally relevant exposure, it may also be a good proxy for any number of other instances of the exposure. For example, if the expression of a particular gene is relatively stable across various tissues, the expression in a specific tissue will likely be a good proxy for expression in other tissues. As such, even if we use expression in that tissue as the exposure, we cannot know if the causal effect β_{XY} is indeed specific to that tissue. Similarly, we also generally do not know other aspects of the exposure such as the dosage, duration and frequency, also limiting the specificity of our conclusions [10, 41, 42].

CONCLUSION

In this Perspective we have outlined how the different assumptions and elements of the data figure into an MR analysis. This outline is not exhaustive, but should provide further insight in how the different components of MR fit together, on both a mathematical and conceptual level. Throughout this paper we have entertained the hypothetical that we know all true associations, focusing specifically on the challenges that remain even in such an idealized scenario. These challenges become substantially harder when having to deal with all the uncertainty in the estimates as well.

As we have shown, causal inference with MR strongly depends on its assumptions. When performing an MR study, it is thus crucial that the validity of these assumptions is examined for each specific analysis, with all alternative scenarios can be carefully considered and ruled out as much as possible. Consequently, performing a reliable MR study requires a considerable investment of time and effort, and access to high quality data for both exposures and outcomes. Despite all its complications however, a well-executed MR study can be a valuable tool in providing greater insight in the relations between our phenotypes. Moreover, the data we have available continues to improve, with more detailed measurements of phenotypes in ever larger biobanks, and rapid innovation in new data and technologies in molecular genetics. With this growth of our data, and our understanding of phenotypes, opportunities for well-designed MR studies will continue to improve.

REFERENCES

1. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol. 2019;2:9.
2. Pearl J. Causal inference in statistics: an overview. Stat Surv. 2009;3:96–146.
3. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23:R89–98.
4. von Hinke Kessler Scholder S, Smith GD, Lawlor DA, Propper C, Windmeijer F. Mendelian randomization: the use of genes in instrumental variable analyses. Health Econ. 2011;20:893–6.
5. Sleiman PMA, Grant SFA. Mendelian randomization in the era of genomewide association studies. Clin Chem. 2010;56:723–8.
6. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Smith GD. Statistical commentary best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. Am J Clin Nutr. 2016;103:965–78.

7. Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. Emerg Themes Epidemiol. 2018;15:1.

8. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations. Wellcome Open Res. 2020;4:186.

9. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. BMJ 2021;375:n2233.

10. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: How to interpret evidence of shared genetic predictors. J Clin Epidemiol. 2016;69:208–16.

11. von Hinke S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Genetic markers as instrumental variables. J Health Econ. 2016;45:131–48.

12. Teumer A. Common methods for performing Mendelian randomization. Front cardiovascular Med. 2018;5:51.

13. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet. 2018;27:R195–208.

14. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun. 2018;9:224.

15. Dai JY, Peters U, Wang X, Kocarnik J, Chang-Claude J, Slattery ML, et al. Diagnostics for pleiotropy in Mendelian randomization studies: global and individual tests for direct effects. Am J Epidemiol. 2018;187:2672–80.

16. 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. 2018;50:693–8.

17. Bowden J, Davey, Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14.

18. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46:1985–98.

19. Burgess S, Zuber V, Gkatzionis A, Foley CN. Modal-based estimation via heterogeneity-penalized weighting: model averaging for consistent and efficient estimation in Mendelian randomization when a plurality of candidate instruments are valid. Int J Epidemiol. 2018;47:1242–54.

20. Qi G, Chatterjee N. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. Nat Commun. 2019;10:1941.

21. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32:377–89.

22. Bucur IG, 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. 2020;29:1081–111.

23. Darrous L, Mounier N, Kutalik Z. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. Genet Genom Med. 2020. http://medrxiv.org/lookup/doi/10.1101/2020.01.27.20018929.

24. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet. 2020;52:740–7.

25. Cho Y, Haycock PC, Sanderson E, Gaunt TR, Zheng J, Morris AP, et al. Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework. Nat Commun. 2020;11:1010.

26. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat Med. 2017;36:4705–18.

27. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? Int J Epidemiol. 2019;48:691–701.

28. Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a trialist?: deconstructing the analogy between Mendelian randomization and randomized trials. Epidemiology. 2017;28:653–9.

29. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls. Epidemiology. 2010;21:383–8.

30. Chen L, Davey Smith G, Harbord RM, Lewis SJ. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. PLoS Med. 2008;5:e52.

31. Van Kippersluis H, Rietveld CA. Pleiotropy-robust Mendelian randomization. Int J Epidemiol. 2018;47:1279–88.

32. Richardson TG, Sanderson E, Elsworth B, Tilling K, Smith GD. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. BMJ 2020;369:m1203.

33. Streeter AJ, Lin NX, Crathorne L, Haasova M, Hyde C, Melzer D, et al. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. J Clin Epidemiol. 2017;87:23–34.

34. Sanderson E, Richardson T, Hemani G, Smith GD. The use of negative control outcomes in Mendelian Randomisation to detect potential population stratification or selection bias. bioRxiv. 2020. https://doi.org/10.1101/2020.06.01.128264.

35. Hughes RA, Davies NM, Davey Smith G, Tilling K. Selection bias when estimating average treatment effects using one-sample instrumental variable analysis. Epidemiology. 2019;30:350–7.

36. Smit RAJ, Trompet S, Dekkers OM, Jukema JW, Le, Cessie S. Survival bias in Mendelian randomization studies: a threat to causal inference. Epidemiology. 2019;30:813–6.

37. Swanson SA. A practical guide to selection bias in instrumental variable analyses. Epidemiology. 2019;30:345–9.

38. Pierce BL, Vanderweele TJ. The effect of non-differential measurement error on bias, precision and power in Mendelian randomization studies. Int J Epidemiol. 2012;41:1383–93.

39. Hemani G, Tilling K, Davey, Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 2017;13:1–22.

40. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. Eur J Epidemiol. 2018;33:947–52.

41. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. BMJ. 2012;345:1–6.

42. Swanson SA, Hernan MA. The challenging interpretation of instrumental variable estimates under monotonicity. Int J Epidemiol. 2018;47:1289–97.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

CdL wrote and revised the paper. The other authors contributed to revision and editing.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41431-022-01038-5.

Correspondence and requests for materials should be addressed to Christiaan de Leeuw.

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

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