doi:10.20944/preprints202101.0035.v1

Understanding the assumptions underlying Mendelian Randomization

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

Affiliations:

[1] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, The Netherlands.

[2] Department of Data Science, Institute for Computing and Information Sciences, Radboud University Nijmegen, The Netherlands

[3] Department of Clinical Genetics, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, The Netherlands.

*Correspondence to: Christiaan de Leeuw Department of Complex Trait Genetics, VU University, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands. Phone: +31 20 598 2832, c.a.de.leeuw@vu.nl

Abstract:

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 strongly 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.

doi:10.20944/preprints202101.0035.v1

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[1,2]. 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 non-experimental methods can be applied to estimate causal relations indirectly[3]. 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[4–6]. MR has become very popular in recent years, with thousands of methodological and applied MR studies published to date[7–10], 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[11–13]. 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[14–16].

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. In order to do so, we will assume 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[15,17].

To serve as a valid instrument for the causal effect of X on Y, there must be an association between G_j and X. Moreover, it must be the case that any association of G_j with Y is mediated by X, as depicted in Figure 1a. In other words, associations of G_j directly with Y or with a confounder C of X

and Y, such as shown in Figure 2a, cannot be present. There is no requirement that G_j has a causal effect itself however; if variant j is in LD with causal variants that are valid instruments, then G_j is a valid instrumental variable as well (Figure 3a).

If we assume the effect sizes of all associations and causal effects to be constant (ie. simple linear relations), we can easily see how this can work to provide the causal effect parameter β_{XY} of the effect of X on Y. Denoting the marginal associations of G_j with X and Y as γ_{Xj} and γ_{Yj} respectively, using the notation from Figure 1a we can express these as $\gamma_{Xj} = \alpha_{Xj}$ and $\gamma_{Yj} = \alpha_{Xj}\beta_{XY} = \gamma_{Xj}\beta_{XY}$. In other words, because the association between G_j and Y is fully mediated by X, it is equal to the causal effect β_{XY} scaled by the association between G_j and X. As such, if we define the ratio of marginal effects $\beta_j = \frac{\gamma_{Yj}}{\gamma_{Xj}}$, it follows that if variant j is a valid instrument then $\beta_j = \frac{\gamma_{Xj}\beta_{XY}}{\gamma_{Xj}} = \beta_{XY}$[17].

We can thus obtain β_{XY} using any genetic variant for which these assumptions hold[18], and all such variants provide the same causal parameter. Indeed, the same applies to any non-genetic variable to which these assumptions apply. Even though in MR the instrumental variables are inherently genetic in nature, it obtains the entire phenotypic causal effect rather than a causal effect specific only to the genetics of X and Y.

This requires that G_j is indeed a valid instrument, which absent further data and analysis must simply be assumed. The a priori plausibility of this assumption varies greatly, depending particularly on the exposure being studied. If for example we have a very proximal endophenotype such as the expression of a particular gene as exposure, and use an exonic variant from that gene for which we have strong experimental evidence of its direct causal effect on the gene's expression, this assumption may be very plausible. But if on the other hand our exposure is an adult behavioural trait with largely unknown biological underpinnings and we only have a weakly associated variant, it probably isn't.

MR also generally depends on some additional assumptions[12,15], which are listed in Table 1. Different methods allow these additional assumptions to be relaxed in various ways so these are not always all required. Making those additional assumptions yields the simplest scenario for MR to tackle. As such, in the next section we will first examine different strategies designed to test and correct for potential violations of the instrumental variable assumptions in the context of these additional assumptions being true. Following that we discuss the role of the additional assumptions and what can happen if they do not hold, in the section *Relaxing the additional assumptions*. An overview of some of the main methods referenced is given in Table 2.

Evaluating instrumental variable assumptions

Using multiple variants

Assuming that the additional assumptions in Table 1 hold, all genetic variants will conform to the scenario in Figure 2a[15,17]. It is composed of reciprocal causal effects β_{XY} and β_{YX} between X and Y, as well as causal effects β_{CX} and β_{CY} on both of a confounder C. In relation to the genetic variant j, it allows for direct associations between G_j and all three variables X, Y and C; associations are direct if they are not mediated by any of the other variables in the figure. Note that the scenario in Figure 2a represents a general model: all the other scenarios in all three figures (except 3c) are a special case of this model, with some of the parameters from Figure 2a set to 0.

To simplify notation all the variables are assumed to be standardized, which for Figure 2a yields $\gamma_{Xj} = \frac{1}{1-\beta_{XY}\beta_{YX}}\left(\alpha_{Xj} + \alpha_{Yj}\beta_{YX} + \alpha_{Cj}(\beta_{CX} + \beta_{CY}\beta_{YX})\right)$ and $\gamma_{Yj} = \frac{1}{1-\beta_{XY}\beta_{YX}}\left(\alpha_{Yj} + \alpha_{Xj}\beta_{XY} + \alpha_{Cj}(\beta_{CY} + \beta_{CX}\beta_{XY})\right)$ for the associations with G_j. From this it is readily apparent that, given the dependence of these terms on multiple variant-specific parameters, the ratio $\beta_j = \frac{\gamma_{Yj}}{\gamma_{Xj}}$ will in the general case be quite specific to each variant as well[12]. This applies to the other scenarios in Figure 2 as well. For example, for Figure 2b we have $\beta_j = \beta_{XY} + \frac{\alpha_{Cj}\beta_{CY}+\alpha_{Yj}}{\gamma_{Xj}}$, simplifying to $\beta_j = \beta_{XY} + \frac{\alpha_{Cj}\beta_{CY}}{\gamma_{Xj}}$ and $\beta_j = \beta_{XY} + \frac{\alpha_{Yj}}{\gamma_{Xj}}$ respectively for Figures 2c and 2d. This is in contrast to variants that are valid instruments (Figure 1a), for which as previously noted the β_j all equal β_{XY}.

It follows that if we have a set of variants, if not all their β_j are the same then at least some of those variants are not valid instruments. As such, if we assume that at least a subset of our variants are valid instruments, we can identify the homogeneous subset of variants that have the same β_j, and obtain β_{XY} from that subset[19].

In practice, this can be accomplished by performing heterogeneity testing, pruning away variants with heterogeneous β_j[20–25]. A variation on this approach can also be used with multiple variants from a single locus[26], to detect the LD-induced pleiotropy scenario in Figure 3b, which is a special case of 2c (see Supplemental Information). A conceptually similar alternative approach is to assume that the valid subset is either the majority or a plurality, and use respectively the median[27] or mode[28–30] of the β_j to obtain β_{XY} (see Supplemental Information).

Such a strategy can identify many variants as invalid instruments to be (implicitly or explicitly) disregarded, but provides no guarantee that the remaining homogeneous subset consists of valid instruments since homogeneity can arise from all three of the scenarios in Figure 1. Unless the scenarios in Figure 1b and 1c can be explicitly ruled out, performing MR using such a homogeneous set of variants therefore still requires the assumption that these variants are valid instruments (Figure 1a).

The two alternative scenarios that need to be ruled out are reverse causation (Figure 1b), with the causal effect operating in the opposite of the hypothesized direction; and the 'mediating confounder' scenario, where the variants are directly associated with a confounder C, which mediates their effects on X and Y (in this scenario, different homogeneous subsets would arise for different confounders).

Distinguishing reverse causation (Figure 1b) from the forward causation scenario in Figure 1a is generally possible. In the notation used here the β_{XY} and β_{YX} represent standardized causal effects; as such β_{XY}^2 and β_{YX}^2, the proportion of variance explained by the causal effects, must both be between 0 and 1. In Figure 1a we have $\beta_j = \beta_{XY}$ for all variants, and thus likewise $0 \leq \beta_j^2 \leq 1$. By contrast, for Figure 1b $\beta_j = \frac{1}{\beta_{YX}}$ for all variants, which means that in this case $\beta_j^2 \geq 1$. By inspection of the β_j^2 value we can therefore rule out one of these two scenarios (except when $\beta_j^2 = 1$).

The mediating confounder scenario in Figure 1c imposes no such constraints on β_j. In this scenario $\beta_j = \frac{\beta_{CY}}{\beta_{CX}}$, or more generally $\beta_j = \frac{\beta_{CY}+\beta_{CX}\beta_{XY}}{\beta_{CX}+\beta_{CY}\beta_{YX}}$ if we allow for direct causal effects between X and Y, and this ratio can take any value. Other means are therefore needed to rule out the mediating confounder scenario.

Instead of using homogeneous subsets as described thus far, we can formulate an alternative strategy based on the observation that, if there is no reciprocal causation ($\beta_{YX} = 0$, Figure 2b), we can express γ_{Yj} as $\gamma_{Yj} = \beta_{XY}\gamma_{Xj} + \delta_j$. This suggests an approach of essentially using a linear regression with γ_{Xj} serving as a predictor for the outcome γ_{Yj}, where all terms not involving β_{XY} are subsumed in the deviation term δ_j: if we assume that γ_{Xj} is (linearly) independent of δ_j, this should have a slope equal to β_{XY} and residuals $\varepsilon_j = \delta_j - b_0$, with b_0 the regression intercept. This is effectively what the commonly used MR-Egger model[31,32] does, with the independence of δ_j following from its InSIDE assumption.

The advantage of such a strategy is that it allows β_{XY} to be computed even if none of the variants are valid instruments, and for the scenario in Figure 2d where the required independence holds, this indeed works. In other scenarios this independence fails however, such as in Figure 2b where $\gamma_{Xj} = \alpha_{Xj} + \alpha_{Cj}\beta_{CX}$ and $\delta_j = \alpha_{Cj}\beta_{CY} + \alpha_{Yj}$. Here, α_{Cj} effectively acts as a confounder in this regression model, biasing the slope away from β_{XY}. The independence assumption is similarly violated in Figures 2c, 2e and 2f.

In practice there are no clear ways of validating this independence assumption. But even if the assumption is true, we would then still face the same issue as the 'homogeneous subset' strategy above: the independence also holds for the reverse causation and mediating confounder scenarios in Figure 1b and 1c (with $\delta_j = 0$), resulting in a slope equal to the β_j expressions given for those scenarios. It is possible to distinguish reverse causation from forward causation with this type of model as well[33]. And moreover, mixture models can be used to more generally account for the possibility of different variants conforming to different scenarios[33,34], such as some acting directly on X and others acting via a confounder C. But distinguishing β_{XY} from confounder effects remains impossible, unless we include additional data that can help do so[12].

Analysing potential confounders

If a putative confounder C is available as a variable, and still assuming the other assumptions in Table 1 hold, evaluating and correcting for that particular C is relatively straightforward. If C is indeed mediating (part of) the associations of G_j with X and Y, adding C as a covariate to compute the conditional associations $\gamma_{Xj|C}$ and $\gamma_{Yj|C}$ will remove the confounding effect. These conditional associations can subsequently be used to perform the MR analysis, assuming no further violations of instrumental variable assumptions are present. Note however that if C is not a confounder, in some cases collider bias may be induced when conditioning on it[35], and as such care should be taken when including variables as covariates (see also Supplemental Information).

In practice however, many known or potential confounders of X and Y will not be available as variables in the GWAS samples. But if a confounder C is mediating associations of j it must itself be associated with G_j, and this can also be verified if results have been published with C as the outcome. Provided that the GWAS was sufficiently well-powered, lack of association of the variant with C is strong evidence that this C is not mediating any genetic effects on X and Y[36]. A similar strategy is to undertake a general lookup of associations for variant j in published GWAS results, to identify known genetic associations with other phenotypes and then evaluate the plausibility of those phenotypes as confounders of X and Y.

Taking this a step further, GWAS results for a possible confounder C can potentially also be used to correct the estimates, using the same principles as the MR-Egger style regression approach. For the model in Figure 2b, the marginal associations are $\gamma_{Xj} = \alpha_{Xj} + \alpha_{Cj}\beta_{CX}$, $\gamma_{Yj} = \alpha_{Xj}\beta_{XY} + \alpha_{Yj} + \alpha_{Cj}(\beta_{CY} + \beta_{CX}\beta_{XY})$ and $\gamma_{Cj} = \alpha_{Cj}$. With multiple variants available, a similar regression of both γ_{Xj} and γ_{Yj} on γ_{Cj} can be performed to obtain the β_{CX} and $\beta_{CY} + \beta_{CX}\beta_{XY}$ terms, then computing corrected associations $\gamma_{Xj|C} = \gamma_{Xj} - \alpha_{Cj}\beta_{CX}$ and $\gamma_{Yj|C} = \gamma_{Yj} - \alpha_{Cj}(\beta_{CY} + \beta_{CX}\beta_{XY}) = \gamma_{Xj|C}\beta_{XY} + \alpha_{Yj}$[37]. Alternatively, this principle can be implemented as a multiple-exposure model, treating C as a second exposure and including α_{Cj} as additional predictor[38,39]. As with conditioning on C directly, similar care must be taken to account for the possibility that C is not a confounder, in which case the conditioning may induce a bias rather than remove it.

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 not considered as a potential confounder by the analyst to begin with, its effects will not have been accounted for. This poses a major challenge, since any confounder variable is itself a phenotype and almost certainly heritable[40], and any variant directly associated with that confounder will also have associations with X and Y mediated by that confounder.

This implies that in practice all (potential) confounders of X and Y 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 X and Y may be specific to a particular context such as a cell type or developmental time period, and measurements of such confounders would therefore need to be specific to that context as well for the above methods to be able to fully correct for them.

Constrained data

More general strategies for validating the instrumental variable assumptions that do not require explicit testing of individual confounders can be found by leveraging natural constraints of data, a prime example of which is RCT. 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. However, confounding that occurs after assignment can be controlled in the experimental design, such as by keeping background conditions at a constant value, counterbalancing factors across groups, and designing matched control group conditions that allow only the intended exposure to differ between the groups. Such measures all constrain potential confounder variables to specific values, preventing confounding.

In an MR context however, the random assignment to a genotype 'group' essentially happens at conception, the exposure occurs at an unknown time possibly many years later, and measurement of the exposure and outcome typically happens even later still. As such, there is a large window of time in which confounding could arise. Yet although MR does not offer any of the experimental control available in RCT[41], well-chosen data with built-in constraints can mimic it to some extent. A clear example of this would be the use of longitudinal data, for either X or Y or both, which would allow the timing of the causally relevant exposure and of the causal effects to be narrowed down much more. If longitudinal measurements of Y are available, conditioning on the value of Y at an earlier time point also blocks any confounder-mediated genetic effects that occurred prior to that time point[42].

Taking more direct inspiration from RCT is the use of additional data from negative control populations[12,43]. A negative control population is one in which X is constrained to a particular value (usually, but not necessarily, 0), but that in other respects matches the population from which the data for the MR analysis was derived (that is, the relations between all relevant variables are otherwise the same). An example of this would be alcohol consumption as the exposure, using a population where people do not drink alcohol due to religious or cultural taboo as control[44]. Note that a negative control population needs to have an actual constraint on the exposure; simply selecting a subset of a population with $X = 0$ does not work, as this would lead to collider bias (see Supplemental Information).

Because in such a control population the exposure does not vary, causal effects by or on X are blocked. For the general model in Figure 2a, in the negative control population the association for Y therefore reduces to $\gamma_{Yj}^{(C)} = \alpha_{Yj} + \alpha_{Cj}\beta_{CY}$, the sum of the possible paths that bypass X. As such, if variant j is a valid instrument then $\gamma_{Yj}^{(C)} = 0$ in the negative control population. Testing $\gamma_{Yj}^{(C)}$ can thus serve to validate the variant as an instrument, provided the control sample affords sufficient power.

This approach can be further extended to directly estimate the part of γ_{Yj} not mediated by X[45], although this only works if there is no reciprocal causation ($\beta_{YX} = 0$) (see Supplemental Information). In this case (Figure 2b), $\gamma_{Yj} = \gamma_{Xj}\beta_{XY} + \alpha_{Yj} + \alpha_{Cj}\beta_{CY}$ and hence $\gamma_{Yj}^{*} = \gamma_{Yj} - \gamma_{Yj}^{(C)} = \gamma_{Xj}\beta_{XY}$. As such, any variant can therefore be used to obtain β_{XY} by means of this corrected γ_{Yj}^{*}, regardless of whether it is a valid instrument. Although potentially quite powerful, this approach is also vulnerable to bias, since if the assumptions of the negative control population (ie. same parameters, and X is fully constrained) fail or $\beta_{YX} \neq 0$, $\gamma_{Yj}^{(C)}$ will not correspond to the non-mediated association between G_j and Y. This is in contrast to merely testing $\gamma_{Yj}^{(C)} = 0$ to determine the validity of variant j as an instrument, which will instead tend to generate false negatives (rejecting valid instruments as invalid) if the negative control population assumptions do not hold.

Further variations on this approach exist as well, including the use of positive and negative control outcomes, which are outcomes for which we already have strong evidence that they respectively are or are not causally influenced by the exposure[15,46]. These can therefore be used to further test the validity of candidate genetic instruments. It should be noted that use of gene-environment interactions has been claimed to provide an alternative way to estimate the corrected γ_{Yj}^{*}[47]. However, in practice this reduces to an MR analysis using a gene-environment interaction term as an instrumental variable rather than G_j itself, requiring the same assumptions and therefore running into the same kinds of problems (see Supplemental Information).

Use of natural constraints such as in longitudinal data and control populations or outcomes has the potential to considerably strengthen support for the instrumental variable assumptions. Strategies like these do require availability of the right data to work, and how difficult such data will be to obtain will vary considerably depending on the exposure and outcome being studied.

Relaxing the additional assumptions

Variable effect sizes

So far we have assumed that all associations and causal effects are all simple linear relations, with effect sizes that are entirely constant and independent of context. In practice however, this assumption may be violated in a number of different ways. One form this can take when γ_{Xj} and γ_{Yj} are obtained from data derived from different populations p and q respectively, with potentially different parameters.

For valid instruments $\beta_j = \frac{1-\beta_{XY}^{(p)}\beta_{YX}^{(p)}}{1-\beta_{XY}^{(q)}\beta_{YX}^{(q)}}\beta_{XY}^{(q)}$ (and equivalent for reverse causation in Figure 1b) , if we assume effect sizes are constant within each population. If there is reciprocal causation in either population this thus introduces a bias, though if this is absent β_j simply reduces to $\beta_{XY}^{(q)}$, the causal effect in population q from which γ_{Yj} was obtained. Differences across the populations in the α_{Xj} for valid instruments will similarly result in the β_j being biased away from β_{XY} for those variants, but this is likely to be accounted for by whatever method is used to deal with heterogeneity of β_j across variants. Somewhat more problematic may be differences in α_{Cj} when using additional GWAS data with C as outcome to test for or correct confounding, as this can bias the correction or reduce power in the test.

Similar issues may also arise even when all data is taken from the same population, if different data sets are subject to different kinds of selection criteria. This implicitly conditions on the variables being selected on, which potentially can both remove mediated and confounded effects of those variables as well as result in collider bias[35,48]. This is also not restricted to explicit selection by the researcher; if for example the outcome is measured specifically in older individuals, this selects for individuals who have survived to that age[49]. Note that selection bias can still be an issue even if it applies equally to all data sets used in the analysis, since it limits the generalizability of the conclusions to the selected subset of the population[50].

Effect sizes may also vary across individuals within a population. This can take the form of non-linear effects of causal variables as well as interactions with other variables. It can also arise as a function of the outcome variable, when causal effects differ in strength depending on the present value of the outcome. A common instance of this latter phenomenon occurs with dichotomous outcome variables[51], for which effects are typically considered linear only on a log-odds or liability scale, and where indeed the binary nature of the variable inherently prohibits effects being linear on the observed scale.

Regardless of their origin, in principle we can approximate all these instances of variable effect sizes by subdividing the population in a set of discrete subpopulations, within each of which the effect sizes are again assumed constant. To give a sense of how this can impact the MR analysis, we will use this approximation to examine the simplest case with only one of the effect size parameter being variable. For each subpopulation p, we will use w_p to denote the size of that subpopulation as a proportion of the whole population.

If the β_{XY} parameter itself is variable, with different causal effect sizes for different subpopulations, then for valid instruments we have $\gamma_{Yj} = \sum_p w_p \gamma_{Xj} \beta_{XY}^{(p)}$ and hence $\beta_j = \sum_p w_p \beta_{XY}^{(p)}$. As such β_j is essentially a weighted mean of the $\beta_{XY}^{(p)}$ (and equivalent for reverse causation in Figure 1b). Although this makes it harder to interpret and generalize, it nevertheless can still be meaningfully interpreted as a sort of average causal effect. Essentially the same thing happens when β_{CX} or β_{CY} are variable: including C (or γ_{Cj}) will only correct for the average confounding effect, but this will generally still be sufficient to remove bias due to that confounder from the β_j.

If the α_{Xj} parameter itself is variable, this will result in $\gamma_{Xj} = \sum_p w_p \alpha_{Xj}^{(p)}$, a weighted mean of the subpopulation associations. Yet $\gamma_{Yj} = \sum_p w_p \alpha_{Xj}^{(p)} \beta_{XY} = \gamma_{Xj}\beta_{XY}$, so because this essentially affects the associations of G_j with X and with Y in the same way it cancels out. More or less the same happens with variability in α_{Yj} and α_{Cj}, with neither necessarily impacting the MR analysis.

What this may seem to suggests is that at least in the simple case of only a single parameter being variable, that variability tends not to strongly impact the MR analysis. But the results above may not hold if γ_{Xj} and γ_{Yj} (or γ_{Cj}, where applicable) are not obtained from the same GWAS cohort. In that case we need to further assume that the proportions w_p are the same across these cohorts. Without this assumption, the above will at best only partially hold true. For example with variable α_{Xj}, for the sum $\sum_p w_p \alpha_{Xj}^{(p)}$ the set of weights w_p implicitly used for γ_{Xj} may differ from those in γ_{Yj}, and hence the relation $\gamma_{Yj} = \gamma_{Xj}\beta_{XY}$ no longer holds.

Unfortunately, variability in parameters is very likely to be accompanied by differences in w_p across cohorts. Unlike linear relations, interactions and non-linear effects are generally sensitive to the distribution of the variables they involve. As such, if variability of parameters is for example caused by an interaction involving a variable D, then simple differences in the mean or variance of D across cohorts will result in different w_P as well, even if all the other parameters are identical. Without further understanding of the source of the variability of the parameters, this would in practice be difficult to correct for. Moreover, scenarios with combinations of multiple variable parameters can be considerably more disruptive still, something that MR analysis using binary phenotypes is thus particularly susceptible to.

Imperfectly observed variables

For the MR analysis, we must also assume that the observed variables (X, Y, and potentially C) for which we computed the associations with G_j are, or are sufficiently good proxies for, the causally relevant variables. This can fail to be the case for a variety of reasons[52,53]. This could be statistical noise, due to measurement error or because the context in which the variable was observed does not sufficiently match that of the causally relevant instance (such as in developmental period, tissue type, or environmental trigger). There can also be more systematic causes. The observed variable may have a complex internal structure, with the causal effect only pertaining to a subtype or subscale of that variable. Similarly, processes such as canalization and behavioural adaptive responses may have amplified or dampened the changes induced by earlier causal effects[16,54,55].

We can represent this as in Figure 3c, where each variable V (with V representing either X, Y or C) is replaced by its causally relevant instance V_c and its observed instance $V_o = \beta_{VO}V_c + \varepsilon_V$ (note that unlike the other variables, these observed instances are not assumed to be standardized). This is a simplified representation, since in practice there may be multiple distinct causally relevant instances, and the relation need not be a simple linear one either (and inherently won't be for dichotomous observed variables). Nevertheless, examining this model can give a sense of the effects imperfect observation can have.

For valid instruments (relative to X_c and Y_c), under this model $\beta_j = \frac{\beta_{YO}}{\beta_{XO}}\beta_{XY}$, showing the attenuation of the causal effect that can arise. In case X_c is fully observed however, $X_o = X_c$ and thus $\beta_{XO} = 1$, and as a result $\beta_j = \beta_{YO}\beta_{XY}$. Although this is still biased relative to β_{XY}, it does have a somewhat meaningful interpretation as the causal effect of X_c on the observed outcome Y_o. In this regard, full observation of the exposure is arguably more important than full observation of the outcome.

Of additional note is that if a variable V_o is only subject to noise relative to its causally relevant instance V_c, ie. $V_o = V_c + \varepsilon_V$, this noise does not bias β_j since in such a case $\beta_{VO} = 1$ and disappears from the equation. However, this only applies if the observed variables remain unstandardized; if they are standardized, $\beta_j = \sqrt{\frac{1+\mathrm{var}(\varepsilon_X)}{1+\mathrm{var}(\varepsilon_Y)}}\beta_{XY}$ and thus any noise will introduce bias in that case.

A further consequence of the attenuation is that it may no longer be possible to distinguish forward and reverse causation[53,56]. Methods that do so directly or indirectly rely on the premise that if the causal direction is correctly specified β_j^2 (or its equivalent, in regression-based approaches like MR-Egger), the variance explained by the causal effect, cannot exceed 1. But since the scaling term $\frac{\beta_{YO}}{\beta_{XO}}$ can take any value, this upper bound ceases to exist when such attenuation is present. Similarly, imperfect observation of a confounder C will also tend to render corrections of the confounding effect only partially effective, as well as reduce power to detect whether C is associated with G_j.

A related issue is that even if the observed X_o is a good proxy for the causally relevant X_c, it may also be a good proxy for any number of other instances of X. 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, further limiting the specificity of our conclusions[16,55,57].

Here again we can see the contrast of MR with RCT[41]. In the latter, the control the researcher has can allow for such specificity to be achieved. This again further suggests using MR with more multivariate and longitudinal measurement of exposures and outcomes, as well as with experimental research on the more proximal effects of variants, which may allow for much more fine-grained conclusions to be drawn.

Conclusion

In this Perspective we have given an outline of how the different assumptions and elements of the data figure into an MR analysis. This outline is by no means exhaustive, but will hopefully provide some 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 we get back to practical reality and need to deal with the uncertainty of our estimates.

As we have shown, the causal inference that MR allows us to perform strongly depends on the assumptions it makes. When performing an MR study, it is thus crucial that the validity of each of these assumptions is examined for each specific analysis, such that all alternative scenarios can be carefully considered and ruled out as much as possible. This is not a challenge unique to MR however, and many of the same issues apply to other methods that have been developed for causal inference using genetic data, such as LCV[58] and GIV[59] (see also Supplemental Information).

Because of this, performing a reliable MR study requires a considerable investment of time and effort, as well as access to high quality data for the exposures and outcomes of interest. Despite all its complications however, when done right MR 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 the rapid innovation in new data and technologies in molecular genetics. And with this growth of our data, and of our understanding of phenotypes, our opportunities for well-designed MR studies will continue to improve as well.

Acknowledgements

This work was funded by The Netherlands Organization for Scientific Research (NWO VICI 453-14-005, 645-000-003, CHiLL 617-001-451).

Author contributions

C.d.L. wrote and revised the manuscript. J.S., I.G.B, T.H. and D.P. contributed to revising and editing the manuscript.

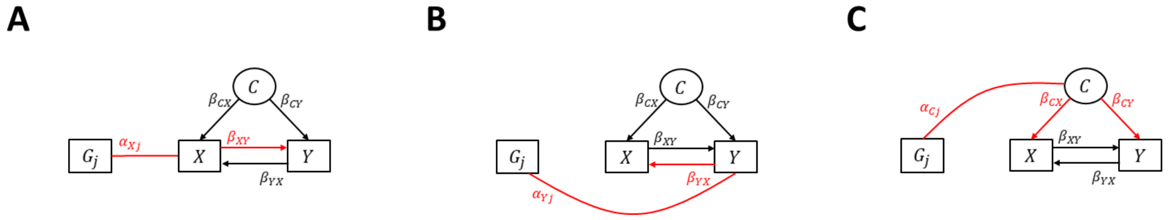

Figure 1. Graphical representation of 'homogeneous' causal scenarios, for a variant j. Each of these three scenarios yield the same $\beta_j = \frac{\gamma_{Yj}}{\gamma_{Xj}}$ ratio for all variants that conform to that scenario. Variables are shown as rectangles or ovals (ovals depict variables that are not (necessarily) observed), with G_j the genotype of j, X the exposure, Y the outcome and C a confounder. Arrows indicate causal effects in the direction of the arrowhead, other lines indicate direct (ie. not mediated by any variable in the graph) associations. Greek letters denote the effect size parameter for each causal effect or association, assuming simple linear relations. For simplicity of notation throughout the paper, all variables are assumed to be standardized, with mean of zero and unit variance. Effects that are required to be non-zero for a scenario are highlighted in red, for **A)** Valid instrument / forward causation scenario, **B)** reverse causation scenario, and **C)** mediating confounder scenario. Note that for the latter mediating confounder scenario, the homogeneity of β_j is specific to each different confounder C.

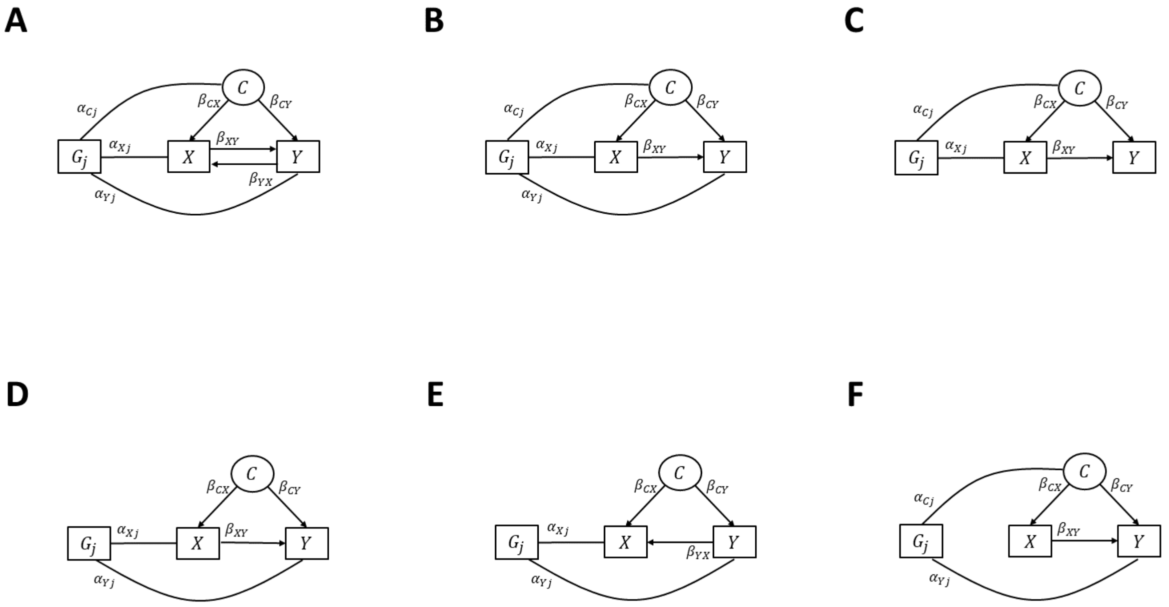

Figure 2. Graphical representation of 'heterogeneous' causal scenarios, for a variant j. All of these scenarios result in different β_j even for variants conforming to the same scenario. Notation and assumptions are the same as in Figure 1. **A)** General model for all scenarios. All other scenarios in Figure 1 and Figure 2 are a special case of this model, equivalent to setting some of the parameters in this model to zero. **B)** Same as A), except assuming no reciprocal causation of Y on X. **C)** Combination of forward causation (Figure 1a) and mediating confounder (Figure 1c) scenarios. **D)** Forward causation scenario with direct pleiotropic effects on Y. **E)** Reverse causation scenario with direct pleiotropic effects on X. **F)** Mediating confounder scenario with direct pleiotropic effects on Y.

doi:10.20944/preprints202101.0035.v1

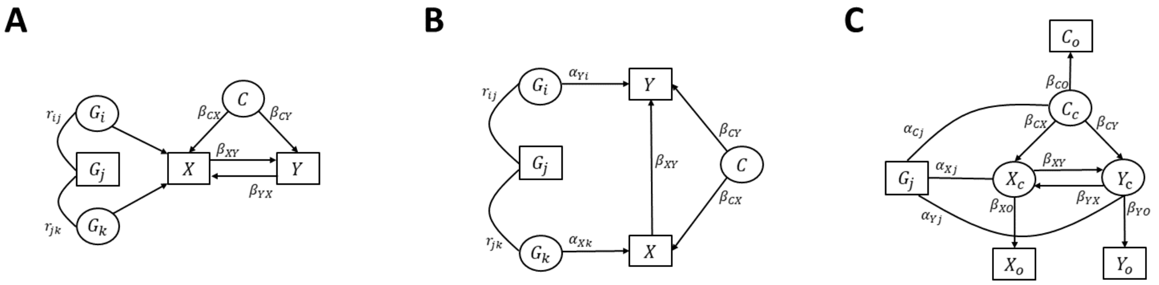

Figure 3. Graphical representation of additional causal scenarios, for a variant *j*. Notation and assumptions are the same as in Figure 1, except that X_o, Y_o and C_o in E) are not assumed to be standardized. **A)** Valid instrument model through LD with causal variants; this is a special case of the scenario in Figure 1a. **B)** LD-induced pleiotropy, with variant j in LD with separate causal variants for X and Y; this is a special case of the scenario in Figure 2d. **C)** Extension of the general model in Figure 2a, distinguishing between the potentially unobserved causally relevant instances of X, Y and C (subscript c), and the instances measured in the data (subscript o).

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 ($\alpha_{Cj} = 0$)
Exclusion restriction	The variant is independent of the outcome given the exposure and all confounders ($\alpha_{Yj} = 0$)
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 non-linearities	Effect sizes for any causal effect or association are not dependent on the value of either of the two variables (as opposed to eg. 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

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

Table 2. Overview of referenced methods

Method	Brief description
Basic multi-variant MR methods	
Two-stage least squares[4]	General instrumental variable analysis model for single-sample MR
IVW mean[4]	Estimates inverse-variance weighted mean of the β_j
Heterogeneity testing	
GSMR[22]	Combination of IVW mean with HEIDI heterogeneity test
GLIDE[24]	Heterogeneity test, using set of simultaneous regression equations
MR-PRESSO	Heterogeneity test, using discrepancy between each variant with IVW estimate based on rest of variants
HEIDI (SMR)[26]	Special application of the HEIDI test for detecting heterogeneity within a locus
Implicit subset MR methods	
Bowden et. al (2016)[27]	Estimates weighted median of the β_j
Hartwig et al. (2017)[28]	Estimates weighted mode of the β_j using empirically smoothed densities
Burgess et al. (2018)[30]	Estimates weighted mode of the β_j using heterogeneity weighted average density of IVW estimates of all subsets of variants
MR-Mix[29]	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[32]	Estimation via weighted linear regression of γ_{Yj} on γ_{Xj}
BayesMR[33]	Bayesian model selection on forward and reverse causation models
CAUSE[34]	Bayesian mixture model allowing a subset of variants to correspond to a different causal scenario
Explicit confounder MR methods	
Multivariable MR-Egger[38]	MR-Egger approach that includes additional γ_{Cj} in the model
MR-TRYX[37]	Large-scale evaluation of potential confounding using GWAS summary statistics database
Negative control population MR methods	
PRMR[45]	Estimates the total component of γ_{Yj} not mediated by X using a negative control population

References

1. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
2. Mills, M. C. & Rahal, C. A scientometric review of genome-wide association studies. *Commun. Biol.* **2**, 9 (2019).
3. Pearl, J. Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96–146 (2009).
4. Burgess, S., Small, D. S. & Thompson, S. G. A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26**, 2333–2355 (2017).
5. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–R98 (2014).
6. Pingault, J. B. *et al.* Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* **19**, 566–580 (2018).
7. Bennett, D. A. & Holmes, M. V. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart* **103**, 1400–1407 (2017).
8. von Hinke Kessler Scholder, S., Smith, G. D., Lawlor, D. A., Propper, C. & Windmeijer, F. Mendelian randomization: the use of genes in instrumental variable analyses. *Health Econ.* **20**, 893–896 (2011).
9. Verduijn, M., Siegerink, B., Jager, K. J., Zoccali, C. & Dekker, F. W. Mendelian randomization: use of genetics to enable causal inference in observational studies. *Nephrol. Dial. Transplant.* **25**, 1394–1398 (2010).
10. Sleiman, P. M. A. & Grant, S. F. A. Mendelian Randomization in the Era of Genomewide Association Studies. *Clin. Chem.* **56**, 723–728 (2010).
11. Haycock, P. C. *et al.* Statistical Commentary Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr* **103**, 965–78 (2016).
12. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27**, R195–R208 (2018).
13. Lousdal, M. L. An introduction to instrumental variable assumptions, validation and estimation. *Emerg. Themes Epidemiol.* **15**, 1 (2018).
14. Harrison, S., Howe, L. & Davies, A. R. *Making sense of Mendelian randomisation and its use in health research A short overview*. (Public Health Wales NHS Trust & Bristol University, 2020).
15. Burgess, S. *et al.* Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res.* **4**, 186 (2020).
16. Burgess, S., Butterworth, A. S. & Thompson, J. R. Beyond Mendelian randomization: How to interpret evidence of shared genetic predictors. *J. Clin. Epidemiol.* **69**, 208–216 (2016).
17. von Hinke, S., Davey Smith, G., Lawlor, D. A., Propper, C. & Windmeijer, F. Genetic markers as instrumental variables. *J. Health Econ.* **45**, 131–148 (2016).
18. Teumer, A. Common Methods for Performing Mendelian Randomization. *Front. Cardiovasc. Med.* **5**, 51 (2018).
19. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–65 (2013).
20. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants. *Epidemiology* **28**, 30–42 (2017).
21. Bowden, J., Hemani, G. & Davey Smith, G. Invited Commentary: Detecting Individual and Global Horizontal Pleiotropy in Mendelian Randomization—A Job for the Humble Heterogeneity Statistic? *Am. J. Epidemiol.* **187**, 2681–2685 (2018).
22. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, (2018).

23. 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).
24. Dai, J. Y. *et al.* Diagnostics for pleiotropy in Mendelian randomization studies: Global and individual tests for direct effects. *Am. J. Epidemiol.* **187**, 2672–2680 (2018).
25. Arellano, M. Sargan's instrumental variables estimation and the generalized method of moments. *J. Bus. Econ. Stat.* **20**, 450–459 (2002).
26. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
27. 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).
28. Hartwig, F. P., Smith, G. D. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
29. Qi, G. & Chatterjee, N. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nat. Commun.* **10**, 1–10 (2019).
30. Burgess, S., Zuber, V., Gkatzionis, A. & Foley, C. N. Modal-based estimation via heterogeneitypenalized weighting: Model averaging for consistent and efficient estimation in Mendelian randomization when a plurality of candidate instruments are valid. *Int. J. Epidemiol.* **47**, 1242–1254 (2018).
31. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–25 (2015).
32. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).
33. 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.* (2019) doi:10.1177/0962280219851817.
34. 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).
35. Gkatzionis, A. & Burgess, S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int. J. Epidemiol.* **48**, 691–701 (2019).
36. Burgess, S. *et al.* Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30**, 543–552 (2015).
37. Cho, Y. *et al.* Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework. *Nat. Commun.* **11**, 1–13 (2020).
38. Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat. Med.* **36**, 4705–4718 (2017).
39. 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).
40. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
41. 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).
42. Streeter, A. J. *et al.* Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *J. Clin. Epidemiol.* **87**, 23–34 (2017).
43. Lipsitch, M., Tchetgen Tchetgen, E. & Cohen, T. Negative Controls. *Epidemiology* **21**, 383–388

(2010).

44. 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).
45. Van Kippersluis, H. & Rietveld, C. A. Pleiotropy-robust Mendelian randomization. *Int. J. Epidemiol.* **47**, 1279–1288 (2018).
46. Sanderson, E., Richardson, T., Hemani, G. & Smith, G. D. The use of negative control outcomes in Mendelian Randomisation to detect potential population stratification or selection bias. *bioRxiv* 2020.06.01.128264 (2020) doi:10.1101/2020.06.01.128264.
47. 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 (2018).
48. 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).
49. Smit, R. A. J., Trompet, S., Dekkers, O. M., Jukema, J. W. & Le Cessie, S. Survival bias in mendelian randomization studies: A threat to causal inference. *Epidemiology* **30**, 813–816 (2019).
50. Swanson, S. A. A Practical Guide to Selection Bias in Instrumental Variable Analyses. *Epidemiology* vol. 30 345–349 (2019).
51. 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).
52. Pierce, B. L. & Vanderweele, T. J. The effect of non-differential measurement error on bias, precision and power in Mendelian randomization studies. *Int. J. Epidemiol.* **41**, 1383–1393 (2012).
53. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, 1–22 (2017).
54. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).
55. Burgess, S., Butterworth, A., Malarstig, A. & Thompson, S. G. Use of Mendelian randomisation to assess potential benefit of clinical intervention. *BMJ* **345**, 1–6 (2012).
56. Burgess, S. & Small, D. S. Predicting the Direction of Causal Effect Based on an Instrumental Variable Analysis: A Cautionary Tale. *J. Causal Inference* **4**, 49–59 (2016).
57. Swanson, S. A. & Hernan, M. A. The challenging interpretation of instrumental variable estimates under monotonicity. *Int. J. Epidemiol.* **47**, 1289–1297 (2018).
58. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018).
59. 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. U. S. A.* **115**, E4970–E4979 (2018).