Richmond, R., & Davey Smith, G. (2021). Mendelian randomization: concepts and scope. In *Combining Human Genetics and Causal Inference to Understand Human Disease and Development* (Cold Spring Harbor perspectives in medicine). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/cshperspect.a040501

Peer reviewed version

Mendelian randomization: concepts and scope

Rebecca C. Richmond[1,2] and George Davey Smith[1,2,3]

[1] MRC Integrative Epidemiology Unit at the University of Bristol, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK

[2] Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK

[3] NIHR Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, UK

Abstract

Mendelian randomization (MR) is a method of studying the causal effects of modifiable exposures (i.e. potential risk factors) on health, social and economic outcomes using genetic variants associated with the specific exposures of interested. MR provides a more robust understanding of the influence of these exposures on outcomes since germline genetic variants are randomly inherited from parents to offspring and, as a result, should not be related to potential confounding factors that influence exposure-outcome associations. The genetic variant can therefore be used as a tool to link the proposed risk factor and outcome, and to estimate this effect with less confounding and bias than conventional epidemiological approaches. We describe the scope of Mendelian randomization, highlighting the range of applications being made possible as genetic datasets and resources become larger and more freely available. We outline the Mendelian randomization approach in detail, covering concepts, assumptions and estimation methods. We cover some common misconceptions, provide strategies for overcoming violation of assumptions, and discuss future prospects for extending the clinical applicability, methodological innovations, robustness and generalizability of MR findings.

Background

Mendelian randomization (MR) was developed as a method to help provide a robust understanding of environmentally modifiable influences on disease [1]. It was proposed to offer a more reliable strategy than conventional observational epidemiological studies which have traditionally been plagued by issues such as confounding (where a common cause of an exposure X and outcome Y may distort the association between X and Y), reverse causation (where Y – or the disease process leading to Y - influences X) and other forms of bias, thus resulting in potentially misleading causal inference [2]. The clearest examples are shown through observational epidemiological studies which have indicated an apparent causal effect that has later failed to be confirmed in large-scale randomized controlled trials (RCTs) [3]. The proposed protective effects of vitamin and antioxidant supplements on cardiovascular disease [4, 5], beta carotene on lung cancer [6, 7], and selenium on prostate cancer [8, 9] are noteworthy examples. Such spurious findings from observational studies have had negative consequences, including the launch of expensive trials based on inadequate evidence, and increased uptake of nutritional supplements in the general population, some of which have subsequently been found to have adverse effects [7, 9].

MR utilizes genetic variants robustly associated with exposures to strengthen inference regarding their potential causal influence on a particular outcome [1, 10]. The online “MR Dictionary” [11] offers a full description and definitions of terminology specific to MR which will be useful to refer to as we elaborate on the concepts and scope of the approach in this paper.

The MR approach draws on Mendel’s laws of segregation and independent assortment, whereby genetic variants are allocated independently of environment and other genetic factors (except those in close physical proximity to the variant of interest, which tend to be inherited together through linkage disequilibrium (LD)) [3]. Based on the premise that the random inheritance of genetic variants from parents to offspring is reflected at a population level, genetic variants can identify groups that differ, on average, by a modifiable exposure. Here, group membership should not be associated with a range of behavioral, social and physiological factors that may confound observational associations [12]. By design, genetic associations should therefore be largely free from confounding and so any difference in outcomes between genetically defined groups can be directly attributed to the exposure.

The association between an outcome and a genetic variant known to proxy a particular risk factor mimics the link between the outcome and the proposed risk factor, and can be used to estimate this relationship with less confounding and bias than conventional epidemiological approaches. Other qualities of (germline) genetic variants which make them useful in causal inference analysis are that they: can be robustly associated with modifiable exposures (i.e. can serve as strong genetic proxies); are fixed at conception and not influenced by disease processes (i.e. are less susceptible to reverse causation); are subject to relatively little measurement error and typically have long-term effects (i.e. are less liable to the underestimation of the exposure-outcome association, referred to as regression dilution bias) [13].

Exposures of interest are typically modifiable and so evidence of causality can – in principle - be used to infer that intervening on an exposure will lead to a change in the outcome under investigation. Making such inference depends upon considering it reasonable to accept the principle of gene-environment equivalence: that perturbation of a phenotype by either a (hypothetical) change in genotype or by environmental change would produce the same downstream effect on an outcome [14-17]. For example, under this assumption, we would anticipate genotypic influence on circulating cholesterol level would lead to the same effect on coronary heart disease (CHD) as would a similar change in cholesterol level induced by dietary influences. While many exposures can be closely proxied by genetic variation, for others it is unlikely that genetic variation will mimic environment exactly, for example in capturing aspects of social deprivation and income [18]. Gene-environment equivalence is a fundamental principle in MR which also brings to the fore the issue of the time-depth of the exposure that is being examined, since genetic variants that influence a phenotype will do so over an extended period. We will come back to the issue of time, discussed at length in the MR literature since its inception [1, 13, 19].

Within a causal inference framework, MR can be implemented as a form of instrumental variable (IV) analysis where the genetic variants serve as proxies or IVs for the modifiable factors of interest (**Figure 1**) [20]. If we suppose X and Y are the exposure and outcome of interest, C is a set of variables that affects X and Y (i.e. potential confounding factors), and U is a further set of variables that affect Y, we can use a further variable G (the genetic variant of interest) as an IV in order to establish the causal effect of X on Y if it satisfies the following assumptions [21]:

1) G is robustly associated with X (“relevance”);
2) G does not share common causes (C and U) with Y (“independence” or “exchangeability”);

3) G affects Y exclusively through its effect on X ("exclusion restriction")

These assumptions are described in more detail in the "Assumptions of Mendelian Randomization and Instrumental Variable analysis" section.

Scope of Mendelian Randomization

Mendelian randomization has been used to:

- appraise the causal relevance of both endogenous (e.g. blood pressure, low-density lipoprotein (LDL) cholesterol) and exogenous exposures (e.g. alcohol, smoking)
- confirm and uncover causal effects for known risk factors of clinical relevance
- establish the causal role of behavioral traits
- evaluate causality in relation to social and economic factors
- assess lifecourse effects
- elucidate intergenerational influences
- characterize difficult to measure environmental exposures
- proxy for modifiers of environmental exposure (e.g. metabolism or detoxification)
- mimic drug targets
- evaluate the role of modifiable mediators between upstream exposures and disease outcomes
- evaluate the effects of genetic liability to a particular disease.

A selection of studies in **Table 1** illustrates how MR has been previously used across a wide variety of contexts.

When the basic principles of MR were initially formalized there were few examples of genetic variants which had robust associations with potentially modifiable exposures, and it was recognized that the future potential of MR would depend upon identifying such associations [1]. There has been very substantial progress in this area. Improvements and cost reductions in array-based genotyping techniques, complemented by DNA sequencing and imputation of information from human genome reference sets, have led to a dramatic increase in our understanding of the genetic contribution to disease risk. Such improvements have also permitted the widespread use of genome-wide association studies (GWAS) which have been successful at detecting replicable associations between common genetic variants and a host of traits in a hypothesis-free approach.

The establishment of genome-wide association study consortia, each focused on investigating different complex traits and diseases, has encouraged numerous population-based studies to contribute genetic data for meta-analysis (**Table 2**). This has in turn increased sample sizes for the discovery and robust replication of GWAS findings. Many of these consortia have also made their GWAS summary data publicly available which, aided by data resources hosting such summary data [22], has catalyzed the development of summary data-based MR studies (described in more detail in "MR Methods").

The recent availability of massive genotyped and phenotyped datasets, including biobank resources (**Table 2**), has added considerably to GWAS efforts. GWAS of phenotypic data from these resources are increasingly performed in an automated fashion, with summary statistics made freely available online (**Figure 2**). Efforts such as these have uncovered a host of genetic variants related to a range of traits, which may leverage greater explanatory power by acting as stronger genetic proxies or instruments in MR [23].

Assumptions of Mendelian Randomization and Instrumental Variable analysis

The key assumption of MR is that of gene-environment equivalence, as discussed above. When using the properties of germline genetic variants to strengthen causal inference, the confidence that a particular modifiable exposure is implicated in the causation of a disease can be enhanced by identifying the direction and magnitude of the effect. This can be estimated through IV analysis. The large majority of MR studies are now implemented within an IV framework, and therefore the IV assumptions are central to MR analysis.

1) Relevance assumption: The genetic variant must be robustly associated with the exposure

The most common method of deriving genetic instruments in recent MR studies is via GWAS, whereby single nucleotide polymorphisms (SNPs) which pass genome-wide significance ($p<5 \times 10^{-8}$) are typically considered for inclusion. However, it is important that the strength of the instrument is tested separately to appraise the relevance assumption, which is often done by means of the proportion of variance explained (r^2) and the related F-statistic, which additionally takes into account the size of the sample under investigation. Increasingly, multiple genetic variants are found to be independently associated with traits investigated in GWAS and these may be combined in genetic risk scores or

through meta-analysis approaches to explain more variation in the trait [23]. This in turn can be used to increase power, obtain more precise causal estimates and minimize risk of weak instrument bias (i.e. uncertainty in the SNP-exposure association which can bias causal estimates) [24].

2) Independence/exchangeability assumption: There are no confounders of the association between the genetic variant and outcome

Since genetic variants are randomized at conception, they should be allocated independent of environmental and other genetic variants excluding those in LD. This means that at a family level, genetic associations should be largely free from conventional confounding. While MR was explicitly introduced in 2003 within a parent-offspring design, data availability did not generally allow use of such designs at the time. It was suggested, however, that population-based studies with appropriate control for population stratification could approximate the parent-offspring design [1, 3]. Concerns about potential violation of this assumption at a population level relate to confounding by ancestry or population stratification, which can influence variation in both allele frequency and disease risk in population(s) being investigated (**Figure 3**). Approaches to limit spurious associations generated because of population groups include use of genetic associations derived from homogeneous populations or with adequate control for population structure e.g. through principal components analysis or linear mixed models [25]. However, the independence assumption can also be violated by dynastic effects (when parental genotypes directly affect offspring phenotypes), or by assortative mating (when individuals select a partner based on a particular phenotype). These biases will likely differ depending on the exposure(s), outcome(s) and population(s) under study.

It is impossible to fully prove the independence assumption in an MR study because, while attempts can be made to account for ancestry and examine how genetic variants relate to measured confounders, associations with unknown confounders cannot be demonstrated. In addition, while previous recommendations have been to assess associations between the genetic instrument and a wide range of potential confounders of the exposure-outcome association [12], where associations are observed, this is unlikely to be due to a direct effect of the confounders on the genotypes and instead could indicate confounding by ancestry (**Figure 3**) or horizontal pleiotropy, as described below (**Figure 4**).

3) Exclusion restriction assumption: The genetic variant should only influence the outcome of interest via the exposure

Biological pleiotropy is the phenomenon whereby a genetic variant influences multiple traits and is a major threat to the exclusion restriction assumption. However, it is important to make the distinction between vertical and horizontal pleiotropy [10, 26]. Vertical (or mediated) pleiotropy occurs where the genetic variant (G) is associated with the outcome (Y) because G affects Y through the exposure (X). This fulfils the exclusion restriction assumption and is the essence of the MR approach. Horizontal (unmediated or biological) pleiotropy occurs when G affects both X and Y but through different pathways. This will often yield biased estimates in MR if a genetic instrument influences the outcome via a mechanism other than the exposure of interest [27]. Such pleiotropy can be direct, as in the path from G to Y (uncorrelated pleiotropy), or can be indirect, e.g. when G affects X and Y through a shared confounder, U (correlated pleiotropy) [28] (**Figure 4**). The latter may occur in cases of mis-specifying the primary phenotype, such as when a genetic variant is used to proxy for a trait secondary to the trait with which it is directly associated.

While it is not possible to prove that the exclusion restriction assumption holds in any MR study, various approaches may be taken to minimize risk:

Use a functional polymorphism for the exposure of interest

One method of ensuring that the genetic variant is unlikely to influence the outcome via another pathway is to use a SNP which has known biological function or is located in a gene which directly codes for the exposure of interest. For example, variants within or near the protein-encoding locus for C-reactive protein (*CRP*) are known to alter serum levels of CRP and are likely to have a predominant influence on any outcomes via this pathway [29].

While single SNPs serve as valid instruments in some situations, in other cases their use is limited if variants do have a pleiotropic effect which cannot be directly estimated. This may be particularly problematic if a variant:

- is associated with multiple biomarkers on separate biological pathways (e.g. genetic variants influencing the branched-chain alpha-ketoacid dehydrogenase (BCKD) enzyme are associated with different branch chain amino acids [30]);

- disrupts the normal function of the exposure (e.g. an *IL6R* variant increases circulating interleukin 6 (IL-6) but reduces IL-6 signaling and therefore decreases risk of coronary heart disease [31]);
- is associated with multiple dependent traits on overlapping pathways, and if those traits have different roles in disease (e.g. *ALDH2* is associated with both alcohol consumption and acetaldehyde level, a known carcinogen, which makes it difficult to disentangle the effects of alcohol and acetaldehyde on risk of esophageal carcinoma [32].

For a detailed description of these scenarios and applied examples see Holmes *et al* [19]. While functional SNP analyses may therefore appear plausible, they can have their drawbacks and many of the sensitivity analyses used for evaluating pleiotropy cannot be applied with a single SNP (see "Methods for assessing and accounting for horizontal pleiotropy"). Instead, it is recommended that associations between the SNP and a wide range of traits are investigated, as described below.

Assess associations between genetic variants and other factors

The presence of associations between genetic variants and other factors may reveal violations to the independence and/or exclusion restriction assumption. A common approach to appraise this is to assess whether the genetic variants used to instrument the exposure (and those variants in LD with the genetic instrument) have been associated with other phenotypes in genome-wide association studies, for example by searching PhenoScanner [33]. While this may highlight genetic variants with horizontal pleiotropy, it can also pick up vertical pleiotropy (e.g. a SNP related to body mass index (BMI) may appear in a GWAS of blood pressure via its influence on BMI). In addition, truly horizontal pleiotropic SNPs may not be detected by this method if the GWAS of the phenotype on the pleiotropic path is absent or underpowered. As such, it is not sufficient to simply exclude the variants which appear in other GWAS as a way to assess the exclusion restriction assumption.

Conduct stratified analysis in a subgroup of the population where the genetic variant is not associated with the exposure of interest

In some instances, conducting a stratified analysis can provide evidence against the possibility of horizontal pleiotropy. When a genetic variant is not related to the exposure of interest in a particular subgroup of the population, this variant should also not be associated with the outcome of interest in this subgroup (given an absence of the association with the exposure). For example, *ALDH2*, coding for

aldehyde dehydrogenase 2, is a common polymorphism in East Asian populations which has been used as a genetic instrument for alcohol consumption [32, 34, 35]. In East Asian populations, where women are much less likely drink to alcohol than men, this polymorphism is not strongly associated with alcohol intake among women [34]. This approach has been used to assess the presence of pleiotropy and evaluate a causal relationship between alcohol consumption and increased blood pressure [34] and risk of vascular disease [35]. For example, if the effects of alcohol consumption on blood pressure and vascular disease are causal, we would expect to find evidence of association between variation in *ALDH2* and the outcomes in East Asian men, but not East Asian women. Any association observed between *ALDH2* and the outcomes in East Asian women, in the absence of alcohol intake, would indicate pleiotropy. Such an approach can be considered a negative control design [36] and additional models build on this approach to detect and adjust for the pleiotropic effects and provide valid estimates in such instances [37, 38] (see "Methods for assessing and accounting for horizontal pleiotropy"). However, genetic variants which are not associated with the exposure in a subgroup of a population may be difficult to identify, and so such direct assessment of pleiotropy is often not as straightforward.

Do not condition on the exposure to assess exclusion restriction

While it may seem intuitive to assess whether statistical adjustment for the exposure leads to attenuation of the gene-outcome relationship, this is not a recommended approach for testing the exclusion restriction assumption. This is because adjusting for the exposure may induce collider bias, where another factor which causes the exposure becomes correlated with the genetic instrument by conditioning on the exposure in this manner [39] (**Figure 5**). It is also worth emphasizing that the stratification approach used in the alcohol consumption example above may also lead to a similar bias if genotypes are not randomly distributed within the strata of exposure. While stratifying by sex is not problematic in this context, since biological sex is not caused by other factors, it is a potential problem in instances where genetic effects are investigated within other subsets of the population, for example if we were to stratify on alcohol drinker status itself [40]. Further to this, bias induced through adjustment for the exposure may be magnified by potential measurement error in the exposure [41].

More advanced methods have been developed to assess violation of the exclusion restriction, including techniques that explicitly model and adjust for pleiotropy, and those that are naturally robust to pleiotropy [26]. These are described in more detail in the "MR Methods" section.

MR Methods

Direct genotype associations

The simplest MR approach to evaluate the presence of a causal relationship is to assess the association between a genetic variant known to influence or modify an exposure and the outcome of interest. However, this does not allow for the magnitude of causal effect to be estimated, which is most often the estimate of interest, especially when considering the translational implications and clinical utility of findings. In addition, multiple pathways can often explain the association between a genetic variant and a particular outcome, so more knowledge of the exposure of interest and its association with the genetic variant is generally required for a valid interpretation [19].

Original applications of one-sample Mendelian randomization

In the pre-GWAS era, most examples of applied MR were conducted using known polymorphisms assayed within one dataset, i.e. where genetic variants, exposure and outcomes of interest obtained from individuals in the same sample (**Figure 6A**). In such a scenario, the causal effect of the exposure on the outcome can typically be estimated using 2-stage least-squares (2SLS) regression [42]. In the first stage, the exposure is regressed on the genetic instrument and in the second stage the outcome is regressed against the predicted values from the first stage. The effect estimate can then be interpreted as the change in the outcome per unit increase in the exposure. The genetic instrument used in one-sample analysis can be a single SNP, multiple SNPs, or a genetic risk score i.e. a summation of risk alleles for each individual which can be unweighted or weighted to give those genetic variants with the strongest effect on the exposure more weight [23].

Studies with more individual-level data may also permit an assessment of associations between genetic variants and confounders of the exposure-outcome relationship in order to investigate the independence and exclusion restriction assumptions. Additional approaches to evaluate violation of the exclusion restriction assumption in one-sample MR include the Sargan test [43] which evaluates heterogeneity of the individual SNP estimates, and IV approaches which can estimate the causal effect in the presence of invalid (e.g. pleiotropic) instruments [44, 45].

Early one-sample MR studies suffered the limitation of low power since few large datasets with relevant genotypic and phenotypic data were available. To counteract this, a number of MR studies were conducted using meta-analysis of causal estimates obtained from independent studies which was

greatly aided by the existence of large genetic consortia [46]. However, the development of two-sample MR analysis has vastly improved the scope of MR applied to large-scale datasets.

Development of two-sample Mendelian randomization

It is possible to use MR to estimate causal effects where genetic associations with the exposure and the outcome have been estimated in different samples (**Figure 6B**). This approach, now known as two-sample MR [24], has greatly increased both the scope and popularity of MR analysis (**Figure 7**). While the initial extended exposition of MR in 2003 [1] included examples of what is now called two-sample MR, the rise in popularity in recent years is attributed to the public availability of GWAS summary data, as well as the development of methods to harmonize and integrate datasets and compute causal estimates when the SNP-exposure and SNP-outcome associations come from different studies [47].

The two-sample approach eliminates the requirement to have access to raw genetic data on individuals within a study and also makes performing MR less time-consuming. In terms of data requirements, all that is needed are the details of the genetic association between the variant(s) and the trait from the exposure GWAS (sample one) and the outcome GWAS (sample two). This typically includes information on the effect and other allele, effect allele frequency, effect estimate and standard error from both GWAS. In addition, the development of web software and code for summary-level data makes MR very straightforward to implement (see "Novel informatic tools"). It also increases the scope of MR analyses, with a wealth of exposure and outcomes available for interrogation which may be infeasible or expensive to measure in the same set of individuals. Furthermore, a series of methods have been developed within this setting in recent years to assess and correct for potential pleiotropy (see "Methods for assessing and accounting for horizontal pleiotropy").

The simplest approach for using summary-level data in an MR framework is to derive a Wald ratio for a single SNP. This is the effect estimate for the SNP-outcome association (from sample 2) divided by the coefficient of the SNP-exposure association (from sample 1), with the standard error of the Wald ratio often approximated by the delta method [48]. In the presence of multiple genetic instruments, a meta-analysis approach (usually inverse-weighted meta-analysis (IVW)) may be used to combine Wald ratio estimates of the causal effect obtained from different SNPs [23]. The point estimates from an IVW MR are equivalent to a weighted linear regression of the SNP-outcome associations on SNP-exposure associations when the intercept is constrained to zero. The effect estimates obtained should also be

equivalent to the effect estimated in 2SLS when sample sizes are large, SNPs independent and there is limited heterogeneity in the Wald ratios.

One- versus two-sample MR

Despite its ease of application, there are various limitations of the two-sample MR which also require consideration. These have been discussed in detail elsewhere [49, 50] and are also summarized in **Table 3.** In part because of these limitations, and also because of the recent availability of large-scale genotyped and phenotyped datasets (**Table 1**), there has been a recent resurgence of one-sample MR. Major benefits of the one-sample MR approach are the flexibility to perform rigorous MR, and the ability to assess the independence and exclusion restriction assumptions through assessment of individual-level confounders.

However, weak instrument bias may threaten the estimation of causal effects in one-sample datasets [51], where uncertainty in the SNP-exposure association could bias the causal estimate (**Table 3**). Importantly, where weak instruments will bias causal estimates in the direction of the null in a two-sample setting, weak instrument bias will be towards the observational association in a one-sample setting [50]. In addition, selection bias due to Winner's curse could lead to biased causal estimates if the genetic variants were discovered in the same sample under investigation. This is a phenomenon which occurs in GWAS by using a p-value cut-off, which can lead to chance over-estimation of the effect size of SNPs with the strongest genetic signals in the GWAS discovery sample [52].

While two-sample methods can be used for one-sample MR analysis [53], these may produce biased estimates and type 1 error rate inflation (i.e. incorrectly rejecting the null hypothesis of no association), something learnt from two-sample MR analyses when genetic consortia have overlapping samples [54]. It is advised that the covariance between the SNP-exposure and SNP-outcome association estimates are taken into account and that external weights be used where possible to minimise the risk of bias. Specifically, genetic variants can be weighted by the magnitude of their association with the exposure in an independent dataset [55], in what could be described as a "one-and-a-half sample MR" design.

Extensions to the basic MR approach

Other methods

A series of developments have been made to extend the application of MR to:

- consider the prevailing direction of causality between two traits (bidirectional MR)
- evaluate intermediates on the causal pathway between exposures and outcomes (two-step, network or mediation MR)
- assess the causal role of closely related traits and to establish independent effects of each (multivariable MR)
- evaluate combined causal effects of risk factors (factorial MR or exposure interactions).

Descriptions, directed acyclic graphs (DAGs) and applied examples for each of these methods are outlined in **Table 4**.

Novel informatic tools

The recent widespread availability of GWAS summary data for a range of traits, with large data repositories and bioinformatic resources for performing MR, provides a powerful and user-friendly way of investigating causal relationships between many different traits [22]. For example, MR-Base is a platform which has retrospectively collected GWAS datasets for more than 20,000 traits, as well as protein-, methylation- and expression-quantitative trait loci (pQTL, mQTL and eQTL) statistics for tens of thousands of molecular markers [47]. Together with its web-based interface which allows the user to explore a range of causal relations, there is an accompanying R package (TwoSampleMR) which allows for LD pruning of genetic instruments in the exposure GWAS, the identification of SNPs and tagging SNPs for each instrument in the outcome GWAS, the careful harmonization of summary statistics between exposure and outcome GWAS, as well as the use of sensitivity analyses to promote evaluation of the MR assumptions.

Methods for assessing and accounting for horizontal pleiotropy

Violation of the exclusion restriction assumption via horizontal pleiotropy is a major threat to the validity of MR analyses and so various methods have been developed in recent years to try to overcome this. These methods: i) can test for potential pleiotropy (e.g. heterogeneity and outlier tests), ii) can directly model and correct for pleiotropy (e.g. MR Egger regression [56] and MR-TRYX [57]) or iii) are "naturally robust" to pleiotropy (e.g. mode and median estimators [58, 59]). These typically use IV estimates as the basis of the sensitivity analyses and can be used to explore how robust MR findings are to the assumption that the genetic instruments used have no horizontal pleiotropic effects.

If the estimate obtained for a causal effect is of a consistent magnitude across multiple independent variants, then pleiotropy is less likely to be a concern. However, often effect estimates are not consistent across independent instruments, with some "outlying" variants having an observed association with the outcome which is substantially different to that expected given their association with the exposure. If the instrument is valid, it should have an effect on the outcome which is proportional to the effect on exposure. Formal tests for examining heterogeneity include the Sargan's test for 2SLS and Cochran's Q statistic, Rucker's Q and likelihood ratio tests in two-sample MR [26, 60]. For detecting outliers, the following approaches can be considered: leave-one-out analysis, Cook's distance, studentized residuals, Q-contribution, and the MR-PRESSO global and outlier tests [27].

Graphical assessment is also helpful for assessing potential pleiotropy. Heterogeneity can be visualized in scatter plots (**Figure 8**), where estimates derived from each genetic variant do not align with the regression line i.e. do not converge to the same causal estimate, or in forest plots where there is clear variation in the causal estimates obtained from each variant. Funnel plot displays of MR estimates of individual genetic variants against their precision will show asymmetry if some variants have unusually strong effects, indicative of pleiotropy. Leave-one-out plots can be used to assess the influence of individual outliers; and the Galbraith Radial plot can be considered in place of the scatter plot for detecting outliers and influential datapoints [61].

While random variation could result in different effects estimated by the individual variants, the presence of heterogeneity in the causal effect estimated by individual SNPs could also indicate horizontal pleiotropy. The simplest method of accounting for this is with the use of a random effects IVW meta-analysis model [23]. However, this approach can only be used where horizontal pleiotropy is *balanced* (i.e. where the random effects have zero weighted mean) [26]. The first method developed for assessing and counteracting the extent of *unbalanced* or *directional* pleiotropy was the application of the Egger regression technique to MR analysis [56]. This approach, first introduced in the context of small study bias within meta-analysis, allows the intercept of the weighted linear regression line of the SNP-outcome on the SNP-exposure association to vary freely (**Figure 8**). Directional pleiotropy can be tested by assessing the extent to which the intercept deviates when it is not constrained to the origin (as in IVW), and the gradient of the line can be used to provide a causal estimate in the presence of directional pleiotropy using the MR Egger approach. It is important to report both estimates.

Additional pleiotropy-robust approaches include the modal and median based estimators which both avoid the contribution of some invalid instruments [58, 62]. Both of these methods can be viewed as implicit outlier correction techniques, since they only allow certain more "reliable" SNPs to contribute to the overall estimate. Use of a weighted approach for both of these methods is typically advocated to maximize statistical power.

Additional methods attempt to account for pleiotropy include: direct outlier removal (e.g. MR-PRESSO [27], generalized summary MR (GSMR) [63], and Cook's distance [64]), outlier penalization (e.g. MR-TRYX [57]); "no-relevance point" approaches including negative controls [65], generalized gene-environment interaction models (MR GxE [37]), pleiotropy-robust MR [66]; and techniques which attempt to directly model pleiotropic pathways including multivariable MR [67], structural equation modelling (SEM) [68] and the Direction of Causation approach (MR-DOC) [68, 69]. These methods are described in more detail elsewhere [26, 70, 71].

It should be emphasized that while these approaches can relax or bypass the exclusion restriction assumption of conventional MR analysis, they in turn come with their own assumptions [26] (**Table 5**). In addition, these approaches are typically less well powered than the IVW approach. As such, these methods for assessing and accounting for pleiotropy should be viewed as sensitivity analyses to conventional approaches such as IVW.

Common misconceptions

With rising popularity of the MR approach which is now becoming a common element of GWAS papers, there are a number of common misconceptions which require debunking in order to ensure that the findings from MR analysis are interpreted appropriately:

"There are three assumptions on which MR relies for estimating causal effects"

While there are three core IV assumptions that apply to many MR studies (relevance, independence and exclusion restriction), additional assumptions are needed to quantify the effect or to consider that the study is informative about effects that may be produced by manipulation of the exposure. The latter is made under the gene-environment equivalence assumption already discussed.

Another assumption in instrumental variable analysis which is relevant for effect estimation is the assumption of homogeneity [72]. This assumption relates to assessing whether the causal effect obtained in MR is uniform across the population, and so represents an average treatment effect (ATE). For example, if investigating the effect of body mass index on cardiovascular disease (CVD), we would assume that a kg/m^2 increase in BMI would elevate risk of CVD irrespective of the person's gender or age. The homogeneity assumption can also be replaced by imposing monotonicity, which assumes that an increase in the number of risk alleles will never lower the likelihood of exposure [73] (e.g. a BMI genetic risk score should not increase BMI for some individuals and decrease it for others). This allows for an estimation of a local average treatment effect (LATE) among those individuals whose exposure level is affected by their genotype. Such assumptions must be made in order for the effect estimate obtained from any MR analysis to be interpreted as the causal effect of the exposure on the outcome. While they cannot be explicitly tested, the assumptions can be interrogated through various means. One possibility is to estimate causal effects in different subpopulations and to evaluate whether the estimated effects differ. Alternatively, as non-homogeneity in the genetic variant – exposure association would lead to non-homogeneity in the genetic variant – outcome association, evaluating the variance of exposure and outcome groups by genotype would provide a test for the presence and degree of violation of this assumption [74]. Large GWAS allow variance to be characterized well, and so may be interrogated to investigate this [75].

An instance where it is difficult to obtain relevant treatment effects from MR estimates is in the presence of binary exposures [76]. As the assumptions of homogeneity and monotonicity are less likely to hold when interpreting causal estimates with binary exposures using MR, it is often simpler to report on the existence and direction (rather than the magnitude) of the causal effect [76]. If these assumptions can be made, there are options for causal estimation with a binary exposure which allow estimates to be converted onto a more clinically meaningful scale. In one-sample data with timed events, it is possible to estimate the causal effect of a binary exposure. However, when conducting two-sample MR, the causal effect estimates should be interpreted as reflecting the effects of the genetic liability to the exposure, rather than the exposure itself [77].

Interpretation of causal effects on binary outcomes is also challenging [78]. While it is empirically possible to calculate causal estimates in a similar manner to continuous variables, for example with use of the log odds scale, the MR effect estimate may only be approximate in the case of a binary outcome [79]. This is because the "non-collapsibility" of odds ratios means that these estimates may not be

constant across strata, and so cannot be used to obtain a precise causal effect [80]. Alternatively, analyses can be conducted on the risk difference scale, which reduces the risk of bias due to non-collapsibility.

As mentioned, while sensitivity analyses may relax some of the core IV assumptions, e.g. exclusion restriction in the case of pleiotropy-robust methods, they impose their own set of (albeit weaker) assumptions (**Table 5**). Two-sample MR also imposes additional assumptions to the one-sample MR approach, including exchangeability of the two samples (i.e. whether they are both drawn from the same underlying population), as well as the assumption that the two samples are non-overlapping [81].

"MR is analogous to a randomized controlled trial"

An analogy has often been made between and MR study and an RCT, where genotypes are said to randomize participants into different levels of exposure or treatment, independent of confounding [82]. In particular, it is this random allocation of genetic variants from parents to offspring which can be viewed as analogous to an RCT [3]. Often this analogy is useful, particularly when it is possible to scale causal effect estimates to that of interventions, for example in the case of (retrospectively) predicting the null effect of selenium on risk of prostate cancer using MR (RR 1.01 (95% CI 0.89, 1.13) per 114 ug/L in MR vs RR 1.04 (95% CI 0.91, 1.19) per 114 ug/L in RCT) [83]. However, the analogy is not perfect since RCTs typically involve interventions of short duration, whereas an individuals' genotype could reflect lifelong exposures, time-dependent or critical period effects, as well as potential developmental compensation that may arise over time [1, 84]. Causal estimates obtained from MR analyses may therefore differ in magnitude to those seen or anticipated in an RCT, although can also be useful in providing added value regarding lifecourse effects, for example of knowledge that cholesterol lowering in earlier life is likely to be important for preventing cardiovascular disease [85]. In addition, in MR analysis conducted at a population- rather than family-level, the analogy is only approximate, as described below.

"Genetic variants are not influenced by confounding factors"

Under Mendel's laws of segregation and independent assortment, it is assumed that genetic variants should be inherited independently of other genetic and environmental factors. While population-level genetic variants are typically much less associated with many potential confounders than directly measured exposures [12], the random inheritance of genetic variants from parents to offspring does not guarantee that genetic variants and confounders will be independent in samples of unrelated

individuals. For example, an obvious violation of this is created due to population stratification which can introduce confounding of genotype-disease associations by factors related to subpopulation group membership within the overall population. This might occur even within groups of relatively homogeneous ancestry, due to underlying substructure [86, 87], or also at the family level, for example due to assortative mating [88]. One potential violation of the second IV assumption is that of "dynastic" or "genetic nurture" effects, where parental genotypes affect the offspring via the environment that parents create for their offspring by affecting the parental phenotype [89]. As a result, a genetic instrument in the offspring will be correlated with the environment created by the parents. One solution to this problem is to perform MR analysis between siblings who have a shared family background and whose genotype differences will not be confounded by parental or family factors (within-family MR) [89, 90]. In the initial presentation of MR it was suggested that the most robust form would be within families [1] and with increasing sources of data for family-based studies, this approach offers potential for elucidating causal effects for those traits which are most likely to be influenced by dynastic effects (e.g. socio-economic factors).

"The exclusion restriction assumption is violated due to pleiotropy"

The exclusion restriction assumption is sometimes referred to as the "no pleiotropy assumption", although it can be violated in a number of other ways, including timing effects, interactions, measurement error, reverse causation, collider bias and LD, as previously described [91].

In particular, the following sources of collider bias may induce spurious associations between a genetic variant and factors other than the exposure of interest that may influence the outcome (**Figure 5**):

- The use of instruments generated from a GWAS which conditions on another phenotype
- Ascertainment bias in case-control studies
- Selection and loss-to-follow up bias in cohort studies
- Survival bias when investigating later-life outcomes
- Evaluation of disease progression in a case-only setting

When there are only moderate effects of a phenotype on selection, bias is generally small [92]. However, where collider bias is likely to exist, it is recommended that sensitivity and simulation studies are carried out to evaluate the extent to which bias might distort estimates [93]. In addition, alternative approaches such as inverse-probability weighting, modeling competing risks, adjusting for index event bias [94, 95], and the use of negative control outcomes can also be considered [96].

Another way in which the exclusion restriction may be violated is when genetic instruments are in LD with other variants which have an effect on the outcome not via the exposure. In this instance, genetic colocalization approaches [97] may be used which attempt to evaluate whether two traits share the same causal variant at a particular locus, and thus contribute to evaluation of whether the exclusion restriction assumption is likely to hold.

"Reverse causation is not an issue for MR"

Since germline genotypes are fixed at conception, they cannot be influenced by reverse causation, and therefore it is often claimed that reverse causation is not an issue for MR. While this is true, a scenario where reverse causation might pose a problem for MR is where there are bidirectional effects between two traits (X and Y) and where the genetic instruments G_X and G_Y are not well characterized [10] (**Table 4**). If trait X influences trait Y, then a GWAS with adequate statistical power will identify a genetic variant with its primary influence of trait X as being associated with trait Y (for example the *CHRNA5* variant related to smoking intensity has been identified at genome-wide significance in relation to lung cancer [98]). This will lead to spurious conclusions if this variant were then used as a genetic instrument for trait Y (e.g. that lung cancer causes smoking), i.e. by mis-specifying the primary phenotype. One way to minimize this problem is to ensure that the two instruments are independent of each other by excluding the genetic variants which they have in common. However, this may also increase the risk of type II error if variants are excluded from the genetic instruments which reflect vertical pleiotropy (e.g. if the *CHRNA5* variant were removed from both the smoking and lung cancer instruments). Alternatively, for a two-sample MR analysis with a binary outcome, it may be possible to investigate the number of "cases" of the outcome in the sample used to run the exposure GWAS. Here a low prevalence of the outcome in the exposure sample would minimize risk of such a reverse signal. Another approach which can be used in this context is the Steiger method applied to MR [99]. This has been developed to investigate whether the genetic variants being used to instrument trait X are more strongly correlated with trait Y than X, in which case they will be excluded from the instrument for X.

One way in which this kind of reverse causation can be leveraged within an MR study is with the notion of "reverse MR" [100]. Here a disease-associated genetic risk score would be expected to associate with causes of the disease e.g. a genetic risk score for lung cancer would be associated with smoking because the *CHRNA5* variant is included in the score. If the outcome of interest cannot plausibly cause the exposure being considered (for example in a subgroup where the outcome is not prevalent, e.g. among

young study participants in the case of lung cancer), then this situation can be used to confirm causal exposures. This principle has been applied to investigate perturbations in proteins and metabolites in relation to later cardiometabolic disease risk [101, 102]. However, alternative scenarios such as pleiotropy, a causal effect of disease liability, or early stages of disease which influence the exposure (e.g. pre-diagnostic cholesterol lowering in cancer [103]) also need to be considered.

Overcoming limitations

Limitations of the MR approach have been discussed extensively elsewhere [1, 3, 10, 13, 49, 50, 91]. While some of the early concerns of the MR approach such as a lack of genetic instruments, horizontal pleiotropy and low power have been ameliorated with larger datasets and novel methods, some limitations remain and new constraints have been recognized.

Lack of reliable polymorphisms for studying modifiable exposures of interest

Genetic instruments extracted from a single gene with a well-understood biological function, and therefore most likely to meet the MR assumptions, are not available for all exposures. While the proliferation of GWAS has increased the availability of genetic variants to use as potential instruments, the role of the variants identified often requires careful consideration to assess their validity in MR. Polygenic influences on most phenotypes imply individual SNPs of small effect size, each with a marginal contribution to the variance explained in a trait. This is both a threat to the exclusion restriction and can lead to problems of weak instruments, unless the variants can be combined into a genetic risk score or applied in large sample sizes.

The use of genetic variants may sometimes lead to counterintuitive results. For example, while it would be expected that longer term IL-6 levels would elevate the risk of CHD [104], genetic variation in the IL-6 receptor which increases circulating IL-6 levels has actually been found to decrease risk of CHD [31]. This can be explained since carriers of this polymorphism exhibiting higher circulating IL-6 levels but reduced membrane-bound IL-6, with reduced IL-6 signaling, which in turn reduces risk of CHD [31]. Similarly, if genetic polymorphisms are associated with multiple aspects of a single trait, for example variation in *CHRNA5* which is associated with different dimensions of smoking behavior (e.g. number of puffs per cigarette, depth of inhalation) [105], this can also lead to problems of the interpretation of causal effects for any particular dimension of the trait.

A further limitation posed by a lack of understanding of genetic variants poses is that of “contamination” of the instrument by variants associated with traits “upstream” of the trait of interest, leading to mis-specification of the primary phenotype. This is a particular concern as GWAS sample sizes increase, since it increases the power to detect genetic variants which act indirectly on the trait of interest. For example, genetic variants with a primary influence on BMI appear among the top hits in GWAS of C-reactive protein [106], but should not be used as instruments for CRP levels. This has already been discussed within the context of a bidirectional relationship, where genetic variants which influence the exposure through reverse cause may be picked up. In addition, it may re-introduce confounding if confounders are picked up as genetic variants for the exposure. For example, genetic variants identified in relation to drinking behavior have been found to be strongly related to socio-economic factors [107]. This may lead to confounding in MR studies if the genetic variants used to instrument the exposure (here drinking behavior) are primarily associated with factors upstream of the exposure (e.g. educational attainment) and may explain the opposite genetic correlations observed between alcohol quantity and intake frequency (which are differentially associated with educational attainment) in relation to various health outcomes in a recent study [108]. Multivariable MR can be used in this instance to estimate the "true” effect of the phenotype being investigated, for example accounting for educational attainment to estimate direct effects of drinking behaviour, but this requires knowing the structure of the underlying phenotypes.

Horizontal pleiotropy

There is a clear trade-off between including more variants in a genetic instrument to maximize variance explained in the exposure and the increased risk of pleiotropy by including more poorly characterized variants. However, the potential advantage of including more variants is the ability to use the suite of approaches already described. These approaches relax the exclusion restriction assumption but each has its own sets of assumptions (**Table 5**). If results are largely consistent across methods which have orthogonal biases [109], there can be more confidence in drawing robust conclusions. Alternatively, in the presence of inconsistent results across methods, a Bayesian model averaging framework may be used as a basis for efficient estimation in the presence of pleiotropy [110].

Lack of independent instruments

While it is not necessary that a genetic variant used to instrument a modifiable exposure of interest in MR should be the causal variant for that trait, it is important to assess whether multiple SNPs used in the instrument are likely to be tagging the same causal variant. This is because including correlated variants will typically lead to erroneous precision in the causal estimate obtained. LD-based clumping or pruning can be used to exclude variants in strong LD, which can ensure independence of the instruments. Another approach is to use a weighted generalized linear regression method which takes into account correlation of multiple non-independent SNPs [111]. This approach is particularly useful when assessing causality of molecular phenotypes (e.g. DNA methylation, gene expression, protein levels), which are often characterized by few independent instruments in *cis* (genetic variants located close to the target locus/gene). A method which is often utilized in conjunction with MR when there is, for example, one independent variant to instrument a molecular trait in *cis*, is the approach of genetic colocalization mentioned previously. Alternatively, the inclusion of *trans* instruments (genetic variants more distal to the target locus/gene) in the MR analysis can be considered, although it is more likely that these variants will violate the exclusion restriction criteria via horizontal pleiotropy.

Need for optimal precision

A failure to recognize the importance of both sample size and instrument strength in MR studies for the detection of expected effects has in the past led to uninformative findings which lack adequate precision to support or reject hypotheses about causal effects. Genotyping in large-scale epidemiological studies as well as the availability of GWAS summary statistics has vastly improved the power of MR studies and has revealed an increasing number of genetic variants which explain a larger proportion of trait variance. Nonetheless, it is always recommended that power calculations for MR studies be conducted a priori [112]. Equivalence testing may also be used to evaluate observed effects within a priori defined equivalence bounds, in order to distinguish effects which are large enough in magnitude to be deemed robust.

It is important to recognize that several of the extensions of the conventional MR approach, such as factorial MR, multivariable MR, sensitivity analyses such as MR-Egger, and within-family MR analyses all suffer from precision constraints which should be taken into consideration. This can be evaluated through additional tests, for example with the Sanderson-Windmeijer conditional F-statistic in the case of multivariable MR which can be used to assess instrument strength for multiple exposures when estimated jointly [113, 114]. Furthermore, precision can be limited in specific contexts, for example

when evaluating intergenerational causal effects which have previously been limited to studies with genetic data available in two generations [115]. In this context, new structural equation model (SEM) approaches have been developed which allow effects of parental exposures on offspring outcomes to be inferred without having to have intergenerational genetic data, and which leverage power from large GWAS summary data in a two-sample approach [68, 69].

Weak instrument bias

Methods to overcome potential weak instrument bias in MR include the use of SIMEX-corrected estimates when the assumption of no measurement error (NOME) cannot be met [116], the use of robust-adjusted profile scores [117], as well as a weak instrument and pleiotropy robust estimation method for use in multivariable MR [114].

Winner's curse

It is recommended that the GWAS discovery sample is independent of the sample(s) used to conduct the MR analysis [49]. Ideally, the genetic variants used as instruments in an MR analysis will also have been replicated in an independent sample in order to further minimize risk of Winner's curse. However, there is a clear trade-off between maximizing sample sizes of GWAS for discovery of genetic variants and avoiding the problem of Winner's curse by retaining a sample for replication and implementation of the MR approach. In the largest datasets, it may be possible to perform a split-sample [118] or jack-knife analysis [119], whereby the dataset is partitioned to avoid problems of sample overlap and Winner's curse.

Canalization and time-varying effects

The notion of canalization or developmental compensation is a potential limitation of MR for which there is no simple empirical assessment [1]. It refers to the buffering of genetic effects during development which may bias MR estimates and vitiate gene-environment equivalence, i.e. that perturbations caused by genotype have the same downstream effects as if they were caused by modifiable exposures. Canalization is a widespread phenomenon in gene knockout studies [1], although it is currently unclear whether the generally small phenotypic differences induced by common functional polymorphisms are sufficient to induce compensatory responses. A related consideration is the often-stated assumption that genetic variants have lifelong effects, which has been used previously to explain large point estimates obtained from MR compared with other epidemiological approaches [120]. There

are clear examples of MR where exposures are time limited, and where canalization is therefore less likely to be an issue. For example, when assessing causal effects of exposures *in utero*, the maternal genetic variants being used as instruments will only have an effect on the offspring via mechanisms during the intrauterine period [115], and when assessing causal effects of exposures which occur predominantly in adulthood (e.g. alcohol consumption, childbearing), the genetic variants will only have an effect after the developmental stage where canalization is most likely to occur.

While it is often difficult to model such time-varying effects in MR, they can bias causal estimates [121] and may have implications for determining optimal timing of interventions. New GWAS studies have started to reveal genetic variants with distinct timing effects [122], which may be leveraged to investigate timing-varying effects in an MR context. For example, a recent multivariable MR study used genetic variants with distinct effects on body size in childhood and adulthood to separate the causal effects of this trait at more than one stage of the lifecourse on risk of chronic disease [123].

Future prospects

While the scope of MR has grown massively in recent years, there are several priority research areas which have not yet been fully evaluated. With increasing methodological and bioinformatic innovation, there is great potential to make progress in these areas. However, subject-specific knowledge, methodological insight and improved reporting of MR findings are required to ensure the robustness and reproducibility of findings.

Extending clinical applicability

Identifying factors underlying disease progression

To date, the majority of GWAS have sought to identify genetic variants associated with risk of disease occurrence. Such variants are informative for disease prevention, but not necessarily for treatment aimed at influencing disease progression since the same genetic factors will not necessarily influence both disease onset and progression of the disease [124, 125]. In 2017, just 8% of genetic association hits in the GWAS Catalog had attempted to identify variants associated with disease progression or severity, and most with modest sample size [125]. Nonetheless, an increasing number of progression GWAS are being carried out, in studies such as the Genetics and Subsequent Coronary Heart Disease consortium (GENIUS-CHD; coronary heart disease, n>270k cases) [126], the Breast Cancer Association Consortium

(BCAC; breast cancer, n>47k cases) [127] and the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome consortium (PRACTICAL; prostate cancer, n>45k cases) [128], allowing for MR studies of progression to be conducted.

However, determining true causal effects on disease progression using MR in case-only datasets is made more challenging because of the issue of collider bias [93, 125]. In particular, when a group of participants are selected based on certain characteristics (e.g. presence of disease), this will introduce a spurious association between independent risk factors influencing selection which will then distort the relationships between each risk factor and disease progression (**Figure 5**). This is a threat to both conventional observational associations and to studies of genetic influences, with confounding being reintroduced which can lead to violation of the MR assumptions. Methods for alleviating such biases are currently in development [93-95]. These attempt to estimate the bias adjustment factor based on estimates for the association of genetic variants with both incidence and progression.

Drug trials

While RCTs remain the gold standard approach for testing the efficacy and safety of a new drug, RCTs can be complemented by MR in terms of prioritizing drug targets, predicting the outcome of trials and optimizing trial design [129, 130]. Human genetic evidence is a strong predictor of drug success [131] and MR studies of proteins and metabolites are becoming fundamental in drug discovery and development [129, 130, 132, 133]. In particular, *cis*-acting variants may serve as genetic proxies for protein drug targets, and selection of such variants may be optimized to evaluate the potential causal relationships between those drug targets and a range of diseases [134-136]. A promising application of MR is in the prioritization of targets for disease prevention, for example revealing the role of PCKS9 inhibitors for reducing LDL cholesterol [137], as well as de-prioritizing interventions based on null results from MR, for example showing that CRP concentration is unlikely to be a causal factor in CHD [138]. MR can also indicate potential side effects of drugs, including the elevated risk of type 2 diabetes among some lipid-lowering drugs [137], which has also been shown in the case of statin trials [139]. It has also been used to identify potential re-purposing opportunities of existing drugs [140]. Most recently, genetic variation in *IL6R* has been associated with lower risk of hospitalization from COVID-19, which is in line with findings from IL6-receptor therapeutic inhibition trials [141]. In addition to predicting the consequences of pharmacotherapy, MR has the potential to be used to optimize trial design, for example in relation to segmenting patients who are most likely to benefit from the drug or giving

insights into the timing of drug initiation [120, 129]. However, this typically requires the availability of prospective datasets of target populations with genetic data available [130].

Scaling up feasibility trials using MR to robustly infer causal effects on clinical endpoints

A novel application of MR has recently been proposed which may enhance the value of feasibility studies of interventions [142]. This approaches uses MR to predict causal effect of these interventions on long-term clinical outcomes via short-term intermediate biomarkers. Feasibility trials are small-scale studies which aim to assess the practicality and acceptability of implementing an intervention in a clinical setting. These are not usually powered to evaluate effects on clinical outcomes and are not typically followed up for enough time to assess long-term outcomes. However, intermediate measures may be collected which can serve as surrogate endpoints in such studies. These measures may lie on the causal pathway to clinical outcomes and MR can be used in this context to appraise the causal effect of those intermediate measures altered by the intervention on long-term outcomes, using a larger study base to the feasibility study in question to bolster power. While these intermediate biomarkers serve as surrogate endpoints, which are well known to have limitations [143], the advantage of using MR in this context is that it is possible to explore both expected and unanticipated effects of manipulating an intermediate trait on a range of outcomes, including relatively rare ones, and with larger sample sizes. Furthermore, the approach may be used to uncover other causal intermediates which could validate choice of surrogate markers for use in future feasibility studies.

Uncovering molecular mechanisms

Building on the success of GWAS and the availability of cost-effective and robust technologies is the scaling up of other 'omic technologies within population health. This is largely concerned with understanding how gene regulatory mechanisms or gene products influence health-related outcomes and is useful for investigating the molecular pathways that may underpin causal effects. In particular, such 'omic measures are influenced by many environmental and endogenous factors and so can be considered as intermediate phenotypes through which causal effects may be investigated [10].

Of particular utility are large-scale 'omic scans for formulating novel hypotheses on biological processes underpinning complex traits and diseases. However, in contrast to germline genetic variation, 'omic signatures are largely phenotypic, and are therefore subject to the same potential problems of confounding and reverse causation which afflict conventional epidemiology [144]. MR is being increasingly applied to elucidate causality for a range of molecular data, including epigenetics,

transcriptomics, gene expression, metabolomics and proteomics [145]. For this, approaches such as two-step, network and multivariable MR are of particular utility for determining whether these markers lie on the causal pathway between risk factors and disease (**Table 4**). Such approaches are being applied in increasingly complex and innovative ways, to consider the causal nature of a large number of molecular markers [146], integrating several types of 'omics data to evaluate molecular pathways [147], as well as considering the tissue-specific nature of causal effects [148].

Increasing ethnic diversity in MR studies

Approximately two thirds of all previous GWAS have been performed in individuals of European ancestry [149]. Differences in allele frequencies and LD patterns between populations threaten the validity of identified genetic variants and therefore transferability of MR findings to other populations [149, 150]. While restricting GWAS and MR analysis to more homogeneous ancestries can help to reduce the threat of population stratification, other approaches such as Bayesian mixture model analysis can be taken to overcome this limitation while ensuring greater diversity in genetic studies [25]. Greater diversity is important as it allows for improved causal inference of risk factors and the clinical translation of genetic findings in other ethnic groups. Furthermore, allowing for diversity in MR studies can help to identify genetic variants which are typically rare in Europeans, where more common variation in other ethnic groups can bolster power in MR for determining causal effects [35, 151]. Some large-scale non-European biobanks are available for genetic analysis (**Table 2**), with a recent GWAS of ~200,000 individuals in Biobank Japan identifying a number of novel loci important for elucidating biology in the East Asian population [152].

Methodological innovations

A number of methodological extensions of the original MR approach have been discussed throughout this review and an increasing number are being developed. In particular, several recently developed whole-genome based approaches, including Genetic Instrumental Variable regression (GIV) [153] and Causal Analysis Using Summary Effect estimates (CAUSE) [28], are seemingly less vulnerable to environmental confounders that are correlated with genes than many of the methods already described (**Figures 3 and 4**). While the previously outlined methods are specifically designed to account for horizontal pleiotropy of the genetic instruments, GIV and CAUSE make use of full GWAS data to also account for other sources of bias. For example, if the primary phenotype has been mis-specified, the Instrument Strength Independent of Direct Effect (InSIDE) assumption of the MR Egger sensitivity

analysis is likely to be violated, and so the CAUSE method may be used as an additional test to determine the presence of correlated pleiotropy in this instance.

While the range of methods now available allows for rigorous analysis and robust causal inference to be made, it can be difficult to navigate the various approaches and appraise their relative strengths and limitations. However, it has been emphasized that the best choice of method can often be context specific [154]. More generally, applying a number of approaches each with orthogonal biases can be helpful in "triangulating" evidence if the estimates obtained from the approaches converge on a similar causal estimate [109].

Automation

The development of bioinformatic platforms and software supports the systematic application of MR [22, 47]. There is scope to automate MR analyses in order to evaluate a multitude of causal relationships in a time-efficient manner. This can aid in the accelerated identification, prioritization and evaluation of intervention targets, for example through phenome-wide association studies (MR-PhEWAS) [155, 156], and the examination of causality in increasingly complex networks with the integration of molecular data [157]. However, limitations of these agnostic approaches include the multiple testing burden imposed as well as the possibility of false positives, which require careful follow-up in terms of evaluating patterns of bias and ensuring robustness of findings to the various assumptions. While machine learning and Bayesian model algorithms have been developed to help select the most appropriate model for evaluation [23, 110, 157-159], users should be careful that the use of automation and data repositories do not trivialize the analysis being conducted and interpretation of results [160].

Improving reproducibility and reporting of MR results

The relative ease at which MR analysis can now be performed can also threaten the design, conduct and reporting of the approach. This may lead to spurious and/or non-reproducible results and may encourage data fishing or the selective cherry-picking of findings, which can lead to study bias in the literature. Guidelines which describe and emphasize the importance of analytical choice considerations and appraise the transparency of MR reporting should help to maintain and improve the quality of MR studies being performed [70, 161, 162]. Code sharing and improved reproducibility of findings, for example emulating recommendations in GWAS to provide independent replication before publication, should also be encouraged.

Conclusions

This paper provides an overarching summary of the Mendelian randomization approach, which uses genetic variants reliably related to modifiable exposures to provide a more robust understanding of the influence of these exposures on disease-relevant outcomes. The development of computational tools and availability of large GWAS datasets has enabled the automation of MR analyses for evaluating a multitude of causal relationships in a time-efficient manner, predominantly via the two-sample MR approach. This has led to the rapid expansion of MR publications and is accelerating the identification, prioritization and evaluation of intervention targets, the detection of novel causal relationships and the integration of molecular data to examine causality in increasingly complex networks. However, as MR is increasingly easy to implement, it may lead to the proliferation of poorly thought-out and conducted studies. It is therefore important that anyone applying the approach is well versed in its assumptions and limitations. We have discussed the current state of the field, highlighting current best practice methodology, methods of assessing the MR assumptions, attempts to overcome potential pitfalls, and some exciting future prospects. Several of the other papers in this collection elaborate on some of the novel methodological approaches, including multivariable MR and the use of MR for assessing mediation [163], polygenic MR methods for assessing pleiotropy [23], as well as within-family MR methods [69]. Other papers describe the application of the approach for extending clinical applicability [129, 130] and uncovering molecular mechanisms [145].

Acknowledgments

We would like to thank Frank Dudbridge, Floriaan Schmidt, Jean-Baptiste Pingault, Fernanda Morales Berstein, Grace Power and Shah Ebrahim for their thoughtful comments on a previous version of the paper. We would also like to thank Fernando Hartwig for his help in producing Figure 7.

References

1. Davey Smith G, Ebrahim S: **'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?** *Int J Epidemiol* 2003, **32**(1):1-22.
2. Davey Smith G, Ebrahim S: **Data dredging, bias, or confounding**. *BMJ* 2002, **325**(7378):1437-1438.
3. Davey Smith G, Holmes MV, Davies NM, Ebrahim S: **Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues**. *Eur J Epidemiol* 2020, **35**(2):99-111.
4. Myung SK, Ju W, Cho B, Oh SW, Park SM, Koo BK, Park BJ, Grp KMAKS: **Efficacy of vitamin and antioxidant supplements in prevention of cardiovascular disease: systematic review and meta-analysis of randomised controlled trials**. *Bmj-Brit Med J* 2013, **346**.
5. Rimm EB, Stampfer MJ, Ascherio A, Giovannucci E, Colditz GA, Willett WC: **Vitamin-E Consumption and the Risk of Coronary Heart-Disease in Men**. *New Engl J Med* 1993, **328**(20):1450-1456.
6. Menkes MS, Comstock GW, Vuilleumier JP, Helsing KJ, Rider AA, Brookmeyer R: **Serum Beta-Carotene, Vitamin-a and Vitamin-E, Selenium, and the Risk of Lung-Cancer**. *New Engl J Med* 1986, **315**(20):1250-1254.
7. Heinonen OP, Huttunen JK, Albanes D, Haapakoski J, Palmgren J, Pietinen P, Pikkarainen J, Rautalahti M, Virtamo J, Edwards BK *et al*: **Effect of Vitamin-E and Beta-Carotene on the Incidence of Lung-Cancer and Other Cancers in Male Smokers**. *New Engl J Med* 1994, **330**(15):1029-1035.
8. Yoshizawa K, Willett WC, Morris SJ, Stampfer MJ, Spiegelman D, Rimm EB, Giovannucci E: **Study of prediagnostic selenium level in toenails and the risk of advanced prostate cancer**. *Jnci-J Natl Cancer I* 1998, **90**(16):1219-1224.
9. Lippman SM, Klein EA, Goodman PJ, Lucia MS, Thompson IM, Ford LG, Parnes HL, Minasian LM, Gaziano JM, Hartline JA *et al*: **Effect of Selenium and Vitamin E on Risk of Prostate Cancer and Other Cancers The Selenium and Vitamin E Cancer Prevention Trial (SELECT)**. *Jama-J Am Med Assoc* 2009, **301**(1):39-51.
10. Davey Smith G, Hemani G: **Mendelian randomization: genetic anchors for causal inference in epidemiological studies**. *Hum Mol Genet* 2014, **23**:R89-R98.
11. Lawlor DA, Wade K, Borges MC, Palmer T, Hartwig FP, Hemani G, Bowden J: **A Mendelian Randomization dictionary: Useful definitions and descriptions for undertaking, understanding and interpreting Mendelian Randomization studies**. 2019.
12. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S: **Clustered environments and randomized genes: A fundamental distinction between conventional and genetic epidemiology**. *Plos Med* 2007, **4**(12):1985-1992.
13. Davey Smith G, Ebrahim S: **Mendelian randomization: prospects, potentials, and limitations**. *Int J Epidemiol* 2004, **33**(1):30-42.
14. West-Eberhard MJ: **Developmental plasticity and evolution**: Oxford University Press; 2003.
15. Ames BN: **Cancer prevention and diet: help from single nucleotide polymorphisms**. *Proc Natl Acad Sci U S A* 1999, **96**(22):12216-12218.
16. Ebrahim S, Davey Smith G: **Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology?** *Hum Genet* 2008, **123**(1):15-33.
17. Davey Smith G: **Epigenesis for epidemiologists: does evo-devo have implications for population health research and practice?** *Int J Epidemiol* 2012, **41**(1):236-247.
18. Hill WD, Hagenaars SP, Marioni RE, Harris SE, Liewald DCM, Davies G, Okbay A, McIntosh AM, Gale CR, Deary IJ: **Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank**. *Curr Biol* 2016, **26**(22):3083-3089.
19. Holmes MV, Ala-Korpela M, Davey Smith G: **Mendelian randomization in cardiometabolic disease: challenges in evaluating causality**. *Nat Rev Cardiol* 2017, **14**(10).
20. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G: **Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology**. *Stat Med* 2008, **27**(8):1133-1163.
21. Hernan MA, Robins JM: **Causal Inference: What If**. Boca Raton: Chapman & Hall/CRC; 2020.
22. Richardson TG, Zheng J, Gaunt TG: **Computational tools for causal inference in genetics**. *Cold Spring Harb Perspect Med* 2020.
23. Dudbridge F: **Polygenic Mendelian Randomization**. *Cold Spring Harb Perspect Med* 2020.

24. Pierce BL, Burgess S: **Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators**. *Am J Epidemiol* 2013, **178**(7):1177-1184.
25. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B *et al*: **Efficient Bayesian mixed-model analysis increases association power in large cohorts**. *Nat Genet* 2015, **47**(3):284-290.
26. Hemani G, Bowden J, Davey Smith G: **Evaluating the potential role of pleiotropy in Mendelian randomization studies**. *Hum Mol Genet* 2018, **27**(R2):R195-R208.
27. Verbanck M, Chen CY, Neale B, Do R: **Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases**. *Nat Genet* 2018, **50**(5):693-698.
28. 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**(7):740-747.
29. Timpson NJ, Nordestgaard BG, Harbord RM, Zacho J, Frayling TM, Tybjaerg-Hansen A, Smith GD: **C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization**. *Int J Obes (Lond)* 2011, **35**(2):300-308.
30. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, Schmidt AF, Imamura F, Stewart ID, Perry JR *et al*: **Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis**. *Plos Med* 2016, **13**(11):e1002179.
31. Swerdlow DI, Holmes MV, Kuchenbaecker KB, Engmann JEL, Shah T, Sofat R, Guo YR, Chung C, Peasey A, Ster RP *et al*: **The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis**. *Lancet* 2012, **379**(9822):1214-1224.
32. Lewis SJ, Davey Smith G: **Alcohol, ALDH2, and esophageal cancer: A meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach**. *Cancer Epidem Biomar* 2005, **14**(8):1967-1971.
33. Staley JR, Blackshaw J, Kamat MA, Ellis S, Young R, Butterworth AS: **PhenoScanner: A Database of Human Genotype-Phenotype Associations**. *Genetic Epidemiology* 2016, **40**(7):664-664.
34. 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**(3):e52.
35. Millwood IY, Walters RG, Mei XW, Guo Y, Yang L, Bian Z, Bennett DA, Chen Y, Dong C, Hu R *et al*: **Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China**. *Lancet* 2019, **393**(10183):1831-1842.
36. Lipsitch M, Tchetgen Tchetgen E, Cohen T: **Negative controls: a tool for detecting confounding and bias in observational studies**. *Epidemiology* 2010, **21**(3):383-388.
37. 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* 2019, **48**(3):702-712.
38. Cho Y, Shin SY, Won S, Relton CL, Davey Smith G, Shin MJ: **Alcohol intake and cardiovascular risk factors: A Mendelian randomisation study**. *Sci Rep* 2015, **5**:18422.
39. Cole SR, Platt RW, Schisterman EF, Chu HT, Westreich D, Richardson D, Poole C: **Illustrating bias due to conditioning on a collider**. *Int J Epidemiol* 2010, **39**(2):417-420.
40. Gage SH, Davey Smith G, Ware JJ, Flint J, Munafo MR: **G = E: What GWAS Can Tell Us about the Environment**. *Plos Genetics* 2016, **12**(2).
41. VanderWeele TJ, Valeri L, Ogburn EL: **The Role of Measurement Error and Misclassification in Mediation Analysis Mediation and Measurement Error**. *Epidemiology* 2012, **23**(4):561-564.
42. Angrist JD, Imbens GW: **2-Stage Least-Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity**. *J Am Stat Assoc* 1995, **90**(430):431-442.
43. Sargan JD: **The Estimation of Economic Relationships Using Instrumental Variables**. *Econometrica* 1958, **26**(3):393-415.
44. Windmeijer F, Farbmacher H, Davies N, Davey Smith G: **On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments**. *J Am Stat Assoc* 2019, **114**(527):1339-1350.
45. Kang H, Zhang AR, Cai TT, Small DS: **Instrumental Variables Estimation With Some Invalid Instruments and its Application to Mendelian Randomization**. *J Am Stat Assoc* 2016, **111**(513):132-144.

46. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, Metrustry S, Cavadino A, Paternoster L, Armstrong LL, De Silva NMG *et al*: **Genetic Evidence for Causal Relationships Between Maternal Obesity-Related Traits and Birth Weight**. *Jama-J Am Med Assoc* 2016, **315**(11):1129-1140.
47. Hemani G, Zheng J, Wade KH, Laurin C, Elsworth B, Burgess S, Bowden J, Langdon R, Tan V, Yarmolinsky J *et al*: **The MR-Base platform supports systematic causal inference across the human phenome**. *Elife* 2018, **7**:e34408.
48. Thomas DC, Lawlor DA, Thompson JR: **Re: Estimation of bias in nongenetic observational studies using "Mendelian triangulation" by Bautista et al**. *Ann Epidemiol* 2007, **17**(7):511-513.
49. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G: **Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies**. *Am J Clin Nutr* 2016, **103**(4):965-978.
50. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, Evans DM, Smith GD: **Recent Developments in Mendelian Randomization Studies**. *Curr Epidemiol Rep* 2017, **4**(4):330-345.
51. Burgess S, Thompson SG: **Bias in causal estimates from Mendelian randomization studies with weak instruments**. *Stat Med* 2011, **30**(11):1312-1323.
52. Garner C: **Upward bias in odds ratio estimates from genome-wide association studies**. *Genet Epidemiol* 2007, **31**(4):288-295.
53. Minelli C, Del Greco MF, van der Plaat DA, Bowden J, Sheehan NA, Thompson J: **The use of two-sample methods for Mendelian randomization analyses on single large datasets**. *Int J Epidemiol* 2021.
54. Burgess S, Davies NM, Thompson SG: **Bias due to participant overlap in two-sample Mendelian randomization**. *Genet Epidemiol* 2016, **40**(7):597-608.
55. Burgess S, Thompson SG: **Use of allele scores as instrumental variables for Mendelian randomization**. *Int J Epidemiol* 2013, **42**(4):1134-1144.
56. Bowden J, Davey Smith G, Burgess S: **Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression**. *Int J Epidemiol* 2015, **44**(2):512-525.
57. Cho Y, Haycock PC, Sanderson E, Gaunt TR, Zheng J, Morris AP, Davey Smith G, Hemani G: **Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework**. *Nat Commun* 2020, **11**(1):1010.
58. Bowden J, Davey Smith G, Haycock PC, Burgess S: **Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator**. *Genetic Epidemiology* 2016, **40**(4):304-314.
59. Hartwig FP, Davey Smith G, Bowden J: **Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption**. *Int J Epidemiol* 2017, **46**(6):1985-1998.
60. 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* 2018, **187**(12):2681-2685.
61. Bowden J, Spiller W, Del Greco MF, Sheehan N, Thompson J, Minelli C, Smith GD: **Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression (vol 47, pg 1264, 2018)**. *Int J Epidemiol* 2018, **47**(6):2100-2100.
62. Hartwig FP, Smith GD, Bowden J: **Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption**. *Int J Epidemiol* 2017, **46**(6):1985-1998.
63. Zhu ZH, Zheng ZL, Zhang FT, Wu Y, Trzaskowski M, Maier R, Robinson MR, McGrath JJ, Visscher PM, Wray NR *et al*: **Causal associations between risk factors and common diseases inferred from GWAS summary data**. *Nat Commun* 2018, **9**.
64. Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Davey Smith G, Timpson NJ: **BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and Understanding Causal Estimates Using Mendelian Randomization**. *Diabetes* 2016, **65**(10):3002-3007.
65. Gage SH, Jones HJ, Burgess S, Bowden J, Davey Smith G, Zammit S, Munafo MR: **Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study**. *Psychological Medicine* 2017, **47**(5):971-980.
66. van Kippersluis H, Rietveld CA: **Pleiotropy-robust Mendelian randomization**. *Int J Epidemiol* 2018, **47**(4):1279-1288.
67. Burgess S, Thompson SG: **Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects**. *Am J Epidemiol* 2015, **181**(4):251-260.

68. Evans DM, Moen GH, Hwang LD, Lawlor DA, Warrington NM: **Elucidating the role of maternal environmental exposures on offspring health and disease using two-sample Mendelian randomization**. *Int J Epidemiol* 2019, **48**(3):861-875.
69. Hwang LD, Davies NM, Warrington NM, Evans DM: **Integrating Family-Based and Mendelian Randomization Designs**. *Cold Spring Harb Perspect Med* 2020.
70. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, Hartwig FP, Holmes MV, Minelli C, Relton C *et al*: **Guidelines for performing Mendelian randomization investigations** *Wellcome Open Res* 2019, **4:186**.
71. Slob EAW, Burgess S: **A comparison of robust Mendelian randomization methods using summary data**. *Genet Epidemiol* 2020, **44**(4):313-329.
72. Swanson SA: **Can We See the Forest for the IVs? Mendelian Randomization Studies with Multiple Genetic Variants**. *Epidemiology* 2017, **28**(1):43-46.
73. Swanson SA, Hernan MA: **The challenging interpretation of instrumental variable estimates under monotonicity**. *Int J Epidemiol* 2018, **47**(4):1289-1297.
74. Mills HL, Higgins JP, Morris RW, Kessler D, Heron J, Wiles N, Davey Smith G, Tilling K: **Detecting heterogeneity of intervention effects using analysis and meta-analysis of differences in variance between arms of a trial**. *Epidemiology* 2020, **(in press)**.
75. Young AI, Wauthier FL, Donnelly P: **Identifying loci affecting trait variability and detecting interactions in genome-wide association studies**. *Nature Genetics* 2018, **50**(11):1608-+.
76. Burgess S, Labrecque JA: **Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates**. *Eur J Epidemiol* 2018, **33**(10):947-952.
77. Richmond RC, Davey Smith G: **Commentary: Orienting causal relationships between two phenotypes using bidirectional Mendelian randomization**. *Int J Epidemiol* 2019, **48**(3):907-911.
78. Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR: **Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses**. *Int J Epidemiol* 2008, **37**(5):1161-1168.
79. Burgess S, Thompson SG, Collaboration CCG: **Methods for meta-analysis of individual participant data from Mendelian randomisation studies with binary outcomes**. *Stat Methods Med Res* 2016, **25**(1):272-293.
80. Greenland S, Robins JM, Pearl J: **Confounding and collapsibility in causal inference**. *Stat Sci* 1999, **14**(1):29-46.
81. Lawlor DA: **Commentary: Two-sample Mendelian randomization: opportunities and challenges**. *Int J Epidemiol* 2016, **45**(3):908-915.
82. Smith GD, Ebrahim S: **What can mendelian randomisation tell us about modifiable behavioural and environmental exposures?** *Bmj-Brit Med J* 2005, **330**(7499):1076-1079.
83. Yarmolinsky J, Bonilla C, Haycock PC, Langdon RJQ, Lotta LA, Langenberg C, Relton CL, Lewis SJ, Evans DM, Smith GD *et al*: **Circulating Selenium and Prostate Cancer Risk: A Mendelian Randomization Analysis**. *Jnci-J Natl Cancer I* 2018, **110**(9):1035-1038.
84. Swanson SA, Tiemeier H, Ikram MA, Hernan MA: **Nature as a Trialist? Deconstructing the Analogy Between Mendelian Randomization and Randomized Trials**. *Epidemiology* 2017, **28**(5):653-659.
85. Ference BA, Julius S, Mahajan N, Levy PD, Williams KA, Flack JM: **Clinical Effect of Naturally Random Allocation to Lower Systolic Blood Pressure Beginning Before the Development of Hypertension**. *Hypertension* 2014, **63**(6):1182-1188.
86. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, Carslake D, Hemani G, Paternoster L, Smith GD *et al*: **Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis**. *Nat Commun* 2019, **10**(1):333.
87. Abdellaoui A, Hugh-Jones D, Yengo L, Kemper KE, Nivard MG, Veul L, Holtz Y, Zietsch BP, Frayling TM, Wray NR *et al*: **Genetic correlates of social stratification in Great Britain**. *Nat Hum Behav* 2019.
88. Hartwig FP, Davies NM, Davey Smith G: **Bias in Mendelian randomization due to assortative mating**. *Genet Epidemiol* 2018, **42**(7):608-620.
89. Brumpton B, Sanderson E, Hartwig FP, Harrison S, Vie GÅ, Cho Y, Howe LD, Hughes A, Boomsma DI, Havdahl A *et al*: **Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases**. *bioRxiv* 2019:602516.

90. Davies NM, Howe LJ, Brumpton B, Havdahl A, Evans DM, Davey Smith G: **Within family Mendelian randomization studies**. *Hum Mol Genet* 2019, **28**(R2):R170-R179.
91. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P: **Methodological challenges in mendelian randomization**. *Epidemiology* 2014, **25**(3):427-435.
92. Gkatzionis A, Burgess S: **Contextualizing selection bias in Mendelian randomization: how bad is it likely to be?** *Int J Epidemiol* 2019, **48**(3):691-701.
93. 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**(3):350-357.
94. Dudbridge F, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, Wain LV, Hingorani AD, Patel RS: **Adjustment for index event bias in genome-wide association studies of subsequent events**. *Nat Commun* 2019, **10**.
95. Mahmoud O, Dudbridge F, Davey Smith G, Munafo M, Tilling K: **Slope-Hunter: A robust method for index-event bias correction in genome-wide association studies of subsequent traits**. *Nat Commun* 2020, **(in press)**.
96. Sanderson E, Richardson TG, Hemani G, Davey Smith G: **The use of negative control outcomes in Mendelian randomization to detect potential population stratification**. *Int J Epidemiol* 2021.
97. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V: **Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics**. *Plos Genetics* 2014, **10**(5).
98. Amos CI, Wu XF, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu XJ, Vijayakrishnan J *et al*: **Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1**. *Nature Genetics* 2008, **40**(5):616-622.
99. Hemani G, Tilling K, Davey Smith G: **Orienting the causal relationship between imprecisely measured traits using GWAS summary data**. *Plos Genetics* 2017, **13**(11).
100. Holmes MV, Davey Smith G: **Can Mendelian Randomization Shift into Reverse Gear?** *Clin Chem* 2019, **65**(3):363-366.
101. Bell JA, Bull CJ, Gunter MJ, Carslake D, Mahajan A, Davey Smith G, Timpson NJ, Vincent EE: **Early Metabolic Features of Genetic Liability to Type 2 Diabetes: Cohort Study With Repeated Metabolomics Across Early Life**. *Diabetes Care* 2020, **43**(7):1537-1545.
102. Ritchie SC, Lambert SA, Arnold M, Teo SM, Lim S, Scepanovic P, Marten J, Zahid S, Chaffin M, Liu Y *et al*: **Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases**. *bioRxiv* 2021:2019.2012.2014.876474.
103. Ahn J, Lim U, Weinstein SJ, Schatzkin A, Hayes RB, Virtamo J, Albanes D: **Prediagnostic total and high-density lipoprotein cholesterol and risk of cancer**. *Cancer Epidemiol Biomarkers Prev* 2009, **18**(11):2814-2821.
104. Danesh J, Kaptoge S, Mann AG, Sarwar N, Wood A, Angleman SB, Wensley F, Higgins JPT, Lennon L, Eiriksdottir G *et al*: **Long-term interleukin-6 levels and subsequent risk of coronary heart disease: Two new prospective studies and a systematic review**. *Plos Med* 2008, **5**(4):600-610.
105. Lassi G, Taylor AE, Timpson NJ, Kenny PJ, Mather RJ, Eisen T, Munafo MR: **The CHRNA5-A3-B4 Gene Cluster and Smoking: From Discovery to Therapeutics**. *Trends Neurosci* 2016, **39**(12):851-861.
106. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, Pellikka N, Wallaschofski H, Kettunen J, Henneman P *et al*: **Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels**. *Circulation* 2011, **123**(7):731-738.
107. Rosoff DB, Clarke TK, Adams MJ, McIntosh AM, Davey Smith G, Jung J, Lohoff FW: **Educational attainment impacts drinking behaviors and risk for alcohol dependence: results from a two-sample Mendelian randomization study with ~780,000 participants**. *Mol Psychiatry* 2019.
108. Marees AT, Smit DJA, Ong JS, MacGregor S, An J, Denys D, Vorspan F, van den Brink W, Derks EM: **Potential influence of socioeconomic status on genetic correlations between alcohol consumption measures and mental health**. *Psychol Med* 2020, **50**(3):484-498.
109. Munafo MR, Higgins JPT, Davey Smith G: **Triangulating Evidence Through the Inclusion of Genetically-Informed Designs**. *Cold Spring Harb Perspect Med* 2020.
110. Shapland CY, Zhao Q, Bowden J: **Profile-likelihood Bayesian model averaging for two-sample summary data Mendelian randomization in the presence of horizontal pleiotropy**. *bioRxiv* 2020:2020.2002.2011.943712.

111. Burgess S, Dudbridge F, Thompson SG: **Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods**. *Stat Med* 2016, **35**(11):1880-1906.
112. Brion MJ, Shakhbazov K, Visscher PM: **Calculating statistical power in Mendelian randomization studies**. *Int J Epidemiol* 2013, **42**(5):1497-1501.
113. Sanderson E, Davey Smith G, Windmeijer F, Bowden J: **An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings**. *Int J Epidemiol* 2019, **48**(3):713-727.
114. Sanderson E, Spiller W, Bowden J: **Testing and Correcting for Weak and Pleiotropic Instruments in Two-Sample Multivariable Mendelian Randomisation**. *bioRxiv* 2020:2020.2004.2002.021980.
115. Lawlor D, Richmond R, Warrington N, McMahon G, Davey Smith G, Bowden J, Evans DM: **Using Mendelian randomization to determine causal effects of maternal pregnancy (intrauterine) exposures on offspring outcomes: Sources of bias and methods for assessing them**. *Wellcome Open Res* 2017, **2**:11.
116. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR: **Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I-2 statistic**. *Int J Epidemiol* 2016, **45**(6):1961-1974.
117. Zhao Q, Wang J, Hemani G, Bowden J, Small DS: **Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score**. *arXiv* 2018.
118. Angrist JD, Krueger AB: **Split-Sample Instrumental Variables Estimates of the Return to Schooling**. *J Bus Econ Stat* 1995, **13**(2):225-235.
119. Angrist JD, Imbens GW, Krueger AB: **Jackknife instrumental variables estimation**. *J Appl Econom* 1999, **14**(1):57-67.
120. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, Kahn J, Afonso L, Williams KA, Sr., Flack JM: **Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis**. *J Am Coll Cardiol* 2012, **60**(25):2631-2639.
121. Labrecque JA, Swanson SA: **Interpretation and Potential Biases of Mendelian Randomization Estimates With Time-Varying Exposures**. *Am J Epidemiol* 2019, **188**(1):231-238.
122. Couto Alves A, De Silva NMG, Karhunen V, Sovio U, Das S, Taal HR, Warrington NM, Lewin AM, Kaakinen M, Cousminer DL *et al*: **GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI**. *Science Advances* 2019, **5**(9):eaaw3095.
123. Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G: **Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study**. *BMJ* 2020, **369**:m1203.
124. Davey Smith G, Paternoster L, Relton C: **When Will Mendelian Randomization Become Relevant for Clinical Practice and Public Health?** *JAMA* 2017, **317**(6):589-591.
125. Paternoster L, Tilling K, Davey Smith G: **Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges**. *PLoS Genet* 2017, **13**(10):e1006944.
126. Patel RS, Tragante V, Schmidt AF, McCubrey RO, Holmes MV, Howe LJ, Direk K, Akerblom A, Leander K, Virani SS *et al*: **Subsequent Event Risk in Individuals With Established Coronary Heart Disease**. *Circ Genom Precis Med* 2019, **12**(4):e002470.
127. Escala-Garcia M, Guo Q, Dork T, Canisius S, Keeman R, Dennis J, Beesley J, Lecarpentier J, Bolla MK, Wang Q *et al*: **Genome-wide association study of germline variants and breast cancer-specific mortality**. *Br J Cancer* 2019, **120**(6):647-657.
128. Szulkin R, Karlsson R, Whitington T, Aly M, Gronberg H, Eeles RA, Easton DF, Kote-Jarai Z, Al Olama AA, Benlloch S *et al*: **Genome-wide association study of prostate cancer-specific survival**. *Cancer Epidemiol Biomarkers Prev* 2015, **24**(11):1796-1800.
129. Ference BA, Holmes MV, Davey Smith G: **Using Mendelian Randomization to Directly Inform and Improve the Design of Randomized Trials**. *Cold Spring Harb Perspect Med* 2020.
130. Schmidt AF, Hingorani AD, Finan C: **Human genomics and drug development**. *Cold Spring Harb Perspect Med* 2021.
131. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J *et al*: **The support of human genetic evidence for approved drug indications**. *Nat Genet* 2015, **47**(8):856-860.

132. Gill D, Georgakis MK, Walker VM, Schmidt AF, Gkatzionis A, Freitag DF, Finan C, Hingorani AD, Howson JMM, Burgess S *et al*: **Mendelian randomization for studying the effects of perturbing drug targets**. *Wellcome Open Res* 2021, **6**:16.
133. Holmes MV, Richardson TG, Ference BA, Davies NM, Davey Smith G: **Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development**. *Nat Rev Cardiol* 2021.
134. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P *et al*: **Genomic atlas of the human plasma proteome**. *Nature* 2018, **558**(7708):73-79.
135. Zheng J, Haberland V, Baird D, Walker V, Haycock P, Gutteridge A, Richardson TG, Staley J, Elsworth B, Burgess S *et al*: **Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases**. *bioRxiv* 2019:627398.
136. Schmidt AF, Finan C, Gordillo-Maranon M, Asselbergs FW, Freitag DF, Patel RS, Tyl B, Chopade S, Faraway R, Zwierzyna M *et al*: **Genetic drug target validation using Mendelian randomisation**. *Nat Commun* 2020, **11**(1):3255.
137. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, Voros S, Giugliano RP, Davey Smith G, Fazio S *et al*: **Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes**. *N Engl J Med* 2016, **375**(22):2144-2153.
138. Wensley F, Gao P, Burgess S, Kaptoge S, Di Angelantonio E, Shah T, Engert JC, Clarke R, Davey-Smith G, Nordestgaard BG *et al*: **Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data**. *BMJ* 2011, **342**:d548.
139. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JE, Shah T, Sofat R, Stender S, Johnson PC, Scott RA *et al*: **HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials**. *Lancet* 2015, **385**(9965):351-361.
140. Li GH, Cheung CL, Au PC, Tan KC, Wong IC, Sham PC: **Positive effects of low LDL-C and statins on bone mineral density: an integrated epidemiological observation analysis and Mendelian randomization study**. *Int J Epidemiol* 2019.
141. Bovijn J, Lindgren CM, Holmes MV: **Genetic variants mimicking therapeutic inhibition of IL-6 receptor signaling and risk of COVID-19**. *Lancet Rheumatol* 2020, **2**(11):e658-e659.
142. Sandu MR, Beynon RA, Richmond RC, Santos Ferreira DL, Hackshaw-McGeagh L, Davey Smith G, Metcalfe C, Lane JA, Martin RM: **Two-step randomisation: applying the results of small feasibility studies of interventions to large-scale Mendelian randomisation studies to robustly infer causal effects on clinical endpoints**. *preprints* 2019.
143. Prentice RL: **Surrogate endpoints in clinical trials: definition and operational criteria**. *Stat Med* 1989, **8**(4):431-440.
144. Relton CL, Davey Smith G: **Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment**. *Plos Med* 2010, **7**(10):e1000356.
145. Porcu E, Sjaarda J, Lepik K, Carmeli C, Darrous L, Sulc J, Mounier N, Kutalik Z: **Causal inference methods to integrate omics and complex traits**. *Cold Spring Harb Perspect Med* 2020.
146. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y *et al*: **Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity**. *Nature* 2017, **541**(7635):81-86.
147. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman AK, Aslibekyan S, Demerath EW, Guan W, Zhi D, Yao C *et al*: **Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach**. *Plos Med* 2017, **14**(1):e1002215.
148. Taylor K, Davey Smith G, Relton CL, Gaunt TR, Richardson TG: **Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization**. *Genome Med* 2019, **11**(1):6.
149. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, Peterson R, Domingue B: **Analysis of polygenic risk score usage and performance in diverse human populations**. *Nat Commun* 2019, **10**(1):3328.
150. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE: **Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations**. *Am J Hum Genet* 2017, **100**(4):635-649.

151. Kang H, Kreuels B, Adjei O, Krumkamp R, May J, Small DS: **The causal effect of malaria on stunting: a Mendelian randomization and matching approach**. *Int J Epidemiol* 2013, **42**(5):1390-1398.
152. Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, Sakaue S, Matoba N, Low SK, Okada Y *et al*: **Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases**. *Nat Genet* 2020.
153. DiPrete TA, Burik CAP, Koellinger PD: **Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data**. *Proc Natl Acad Sci U S A* 2018, **115**(22):E4970-E4979.
154. Koellinger PD, de Vlaming R: **Mendelian randomization: the challenge of unobserved environmental confounds**. *Int J Epidemiol* 2019, **48**(3):665-671.
155. Richardson TG, Harrison S, Hemani G, Davey Smith G: **An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome**. *Elife* 2019, **8**.
156. Millard LA, Davies NM, Timpson NJ, Tilling K, Flach PA, Davey Smith G: **MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization**. *Sci Rep* 2015, **5**:16645.
157. Hemani G, Bowden J, Haycock P, Zheng J, Davis O, Flach P, Gaunt T, Smith GD: **Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome**. *bioRxiv* 2017:173682.
158. Howey R, Shin S-Y, Relton C, Davey Smith G, Cordell HJ: **Bayesian network analysis incorporating genetic anchors complements conventional Mendelian randomization approaches for exploratory analysis of causal relationships in complex data**. *PLoS Genet* 2020, **16**(e1008198).
159. Zuber V, Colijn JM, Klaver C, Burgess S: **Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization**. *Nat Commun* 2020, **11**(1):29.
160. Burgess S, Davey Smith G: **How humans can contribute to Mendelian randomization analyses**. *Int J Epidemiol* 2019, **48**(3):661-664.
161. Davey Smith G, Davies NM, Dimou N, Egger M, Gallo V, Golub R, Higgins JPT, Langenberg C, Loder EW, Richards JB *et al*: **STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies**. *peerj* 2019.
162. Davies NM, Holmes MV, Davey Smith G: **Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians**. *BMJ* 2018, **362**:k601.
163. Sanderson E: **Multivariable Mendelian Randomization and Mediation**. *Cold Spring Harb Perspect Med* 2020.