The use of Mendelian randomisation to identify causal cancer risk factors: promise and limitations

The use of Mendelian randomisation to identify causal cancer risk factors: promise and limitations

Harvinder Gala* and Ian Tomlinson*

Cancer Research UK Edinburgh Centre, Institute of Genetics & Molecular Medicine, The University of Edinburgh, Edinburgh, UK

Correspondence to: H Gala or I Tomlinson, Cancer Research UK Edinburgh Centre, Institute of Genetics & Molecular Medicine, Western General Hospital, The University of Edinburgh, Crewe Road South, Edinburgh EH4 2XR, UK. E-mail: h.gala@sms.ed.ac.uk (Gala) or E-mail: ian.tomlinson@igmm.ed.ac.uk (Tomlinson)

Abstract

The use of observational analyses, such as classical epidemiological studies or randomised controlled trials (RCTs), to infer causality in cancer may be problematic due to both ethical reasons and technical issues, such as confounding variables and reverse causation. Mendelian randomisation (MR) is an epidemiological technique that uses genetic variants as proxies for exposures in an attempt to determine whether there is a causal link between an exposure and an outcome. Given that genetic variants are randomly assigned during meiosis according to Mendel's first and second laws of heritability, MR may be thought of as a 'natural' RCT and is therefore less vulnerable to the aforementioned problems. MR has the potential to help identify new, and validate or disprove previously implicated, modifiable risk factors in cancer, but it is not without limitations. This review provides a brief description of the history and principles of MR, as well as a guide to basic MR methodology. The bulk of the review then examines various limitations of MR in more detail, discussing some of the proposed solutions to these problems. The review ends with a brief section detailing the practical implementation of MR, with examples of its use in the study of cancer, and an assessment of its utility in identifying cancer predisposition traits.

© 2020 The Authors. *The Journal of Pathology* published by John Wiley & Sons Ltd on behalf of Pathological Society of Great Britain and Ireland.

Introduction

One of the many difficulties facing those studying complex diseases such as cancer is the determination of causal risk factors. Cancer appears to be the result of both inherited (genetic) and environmental factors, and highlighting the true causes is critical for the development of appropriate preventive and therapeutic agents. The identification of genetic variants that increase the risk of cancer is ongoing and has sped up greatly due to advances in technologies such as next-generation sequencing. However, the accurate identification of non-genetic risk factors in cancer remains a challenge for a number of reasons. Although many epidemiological techniques may be used to identify associations between traits and cancer, establishing causality is a much more difficult task. Therefore, the development of techniques that may be used to confirm causal links between genetic variation and cancer is extremely important.

For a long time, the gold standard method of inferring causality, and hence suitability of disease interventions, came from randomised controlled trials (RCTs). There have been instances where the proposed cause for an effect from observational studies has later been found to be incorrect from follow-up RCTs, for example, the formerly recommended beta-carotene in cancer prevention [1–6]. However, RCTs may not be able detect exposures that occur over a long period because they have relatively short follow-up times. There may also be ethical and financial reasons as to why RCTs are not a viable means of determining causality [7].

Because genetic associations with cancer risk will not suffer from problems of reverse causation and are unlikely to be affected by other confounders, the use of genetics to infer causality of non-genetic risk factors is attractive and feasible if those risk factors have some genetic basis. This is the rationale for using the technique of Mendelian randomisation (MR) to determine risk factor causality.

Mendelian randomisation (MR)

MR is an epidemiological technique that has been developed as a means of not only avoiding the pitfalls classically associated with RCTs, such as confounding variables, but also examining causal factors for phenotypes that would not be appropriate for RCTs, for instance, height. It is based on the fact that genetic variants—in practice mostly common polymorphisms—are randomly assigned during meiosis according to Mendel's first and second laws of inheritance, thus mitigating the problem of reverse causation.

An early example of the MR concept came from Katan in 1986; he described the problem of determining whether the association between low serum cholesterol and cancer was actually causal, or resulted from the presence of confounding variables such as diet and smoking, or from reverse causation. To solve this, he used the fact that the apolipoprotein E (APOE) gene is polymorphic, with the different alleles encoding isoforms with varying potency with regard to clearing cholesterol from the plasma. Katan concluded that if low serum cholesterol did in fact cause cancer, members of the population with the potent APOE isoforms ought to have a higher frequency of cancer; a distribution that differs from the aforementioned would imply that there is no relationship between low serum cholesterol and cancer [8]. A study performed in 2009 used MR in an attempt to evaluate this relationship and found no increased risk of cancer between groups of elderly patients categorised by APOE genotype [9].

A similar proposal aiming to make an unbiased assessment of allogeneic bone marrow transplants (BMTs) came from Gray and Wheatley in 1991 and is where the term 'Mendelian randomisation' originated. Gray and Wheatley noted that a comparison between allogeneic BMTs and chemotherapy in the treatment of leukaemia was problematic; in cases where a human leukocyte antigen (HLA)–matched donor was available, it would be unethical to withhold BMT from a patient [10]. Their solution was to compare the survival of patients receiving BMTs from HLA-compatible siblings against non–HLA-compatible siblings: Because HLA status is assigned at conception, that is, before the onset of disease, selection bias (discussed below) is avoided [10]. Subsequent trials in patients with acute myeloid leukaemia (AML) testing Gray and Wheatley's proposal have indeed found that HLA compatibility in allogeneic BMTs reduced relapse and may provide a survival advantage [11,12].

MR involves the use of *instrumental variables* (IVs) in the form of germline variants, usually single nucleotide polymorphisms (SNPs). These are used as proxies for *exposures* (or *intermediate traits*) in order to establish a causal link between an exposure and an *outcome*. In Katan's example, the APOE SNP is an IV, cholesterol is the exposure, and cancer is the outcome. In its simplest form, MR concludes that the exposure is causal if its association with the outcome is statistically significant and can be explained entirely by the genetic variant's two associations: (1) with the exposure and (2) with the outcome. MR relies on a number of assumptions for it to be accurate [6,7]. The rationale underlying MR and the required IV assumptions may be visualised using a directed acyclic graph (DAG) (Figure 1) and are as follows:

i. The IVs (the SNPs being used) should be clearly and quantifiably linked to the exposure(s) in question.
ii. The IVs should not be linked in any way to any confounding variables.
iii. The IVs should be linked to the outcome only through the exposure(s) in question.

To estimate a causal effect with IV analysis, additional assumptions are required; one such assumption is that [13]:

iv. The associations are linear and not affected by statistical interactions.
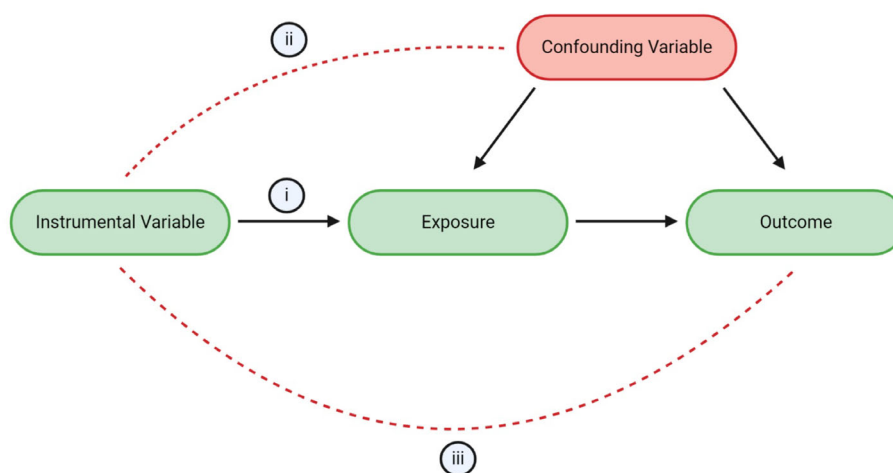

Figure 1. Directed acyclic graph (DAG) depicting MR principles and underlying IV assumptions (i–iii). Black arrows indicate causal links, and red-dashed lines indicate potential violations of MR assumptions. Created using BioRender.com.

Overview of MR methodology

The 'traditional' way of performing MR is the one-sample method and involves the acquisition of SNPs, exposures, and outcomes all from a single data set [14]. However, very few data sets are large enough for one-sample MR to be conducted with sufficient power to infer causal relationships [7]. To increase statistical power, and thereby ameliorate this issue, a method called two-sample MR has more recently become prevalent; this involves obtaining SNP and exposure data from one data set, and SNP and outcome data from another data set, with the two data sets ideally being very similar in terms of their risk factors for the outcome [15].

The initial step in MR involves the estimation of the exposure using valid IVs; the effect size (or beta), standard error (SE), and effect/other alleles are needed for each SNP, with analogous information for the outcome also required [16]. For analysis that involves different genomic regions (polygenic analysis), IVs may be chosen using either a biological or a statistical approach. The biological approach involves choosing IVs that have been linked to the exposure from previous studies, whereas the statistical approach involves the inclusion of all IVs below a genome-wide significance threshold (usually those with a P-value $<5 \times 10^{-8}$) [17]. The outcome is then regressed on the exposure to give the causal effect estimate [18]. The most common type of MR analysis gives higher weighting to SNPs that have smaller SEs in the IV-outcome regression and is called the inverse-variance weighting (IVW) method [16,19].

The most straightforward way of performing MR is called the ratio of coefficients or Wald method. Here, the causal effect is triangulated by dividing the coefficients of regression of the outcome on the IV by the regression of the exposure on the IV [20,21]. The ratio of coefficients method may be performed using two-sample MR and if the outcome is dichotomous (*e.g.* disease versus no disease), but requires that only one IV be used [18,21]. This method requires only the IV-exposure and IV-outcome regression coefficients and can therefore be performed using summary-level data, without the need for individual-level data [21].

Another method of performing MR analysis is called the two-stage least-squares (2SLS) method. 2SLS, as the name implies, involves two stages of regression: The first is from the IVs to the exposure, and the second is from the exposure to the outcome [21]. These two regressions are performed in the same model, and so the covariation of the IVs and the exposure must be taken into account to obtain an accurate SE [18]. However, this method requires individual-level data and becomes biased when at least one invalid IV is used [22]. The control function estimator is a suggested adaptation of the 2SLS approach, where the residuals of the first-stage regression, which correspond to the effects of confounding variables on the exposure-outcome coefficient, are included in the second-stage regression;

[18,21]. This allows for the effects of any confounding variables on the outcome to be controlled for.

The output for MR analysis gives (1) a P-value that represents the probability that the trait being used as an exposure is causal for an outcome, with a P-value <0.05 generally considered to be statistically significant and (2) odds ratios (ORs) that quantify the effect of an exposure on the outcome (for example, OR = 1.10 means that the outcome is 10% more likely to occur with the exposure compared to without it), as well as confidence intervals usually set at 95%, with smaller intervals indicative of greater OR precision [23]. For all of the aforementioned methods, the estimate of the causal effect may be thought of as the change in outcome per unit change in exposure [21].

A technique called multivariable MR is an extension of MR whereby multiple IVs that affect multiple exposures are assessed for their causal effect on an outcome simultaneously [24]. This method can account for measured horizontal pleiotropy (discussed below) and has recently been adapted for high-throughput experiments [25]. In the context of cancer, obtaining results from MR that suggest there are modifiable causal risk factors may allow for risk minimisation, if not prevention.

Limitations of MR

In spite of the fact that MR has been shown to be a useful tool in epidemiology, particularly with regard to mitigating reverse causation and confounding variables, there are several limitations to be considered:

Horizontal pleiotropy

Horizontal pleiotropy is the term used to describe when an IV is linked to the outcome in a way that does not involve the exposure, thus violating the third IV assumption (Figure 2). The violation of this assumption, also known as the 'exclusion restriction criterion', can severely reduce the accuracy of MR, resulting in an incorrect quantification of causality, reduced statistical power, and type I errors [26].

The problem of horizontal pleiotropy appears to be very difficult to circumvent, as it has been found to be abundant in complex human diseases from genome-wide association studies (GWAS) [27,28]. This has led to conflicting opinions on the utility of MR, with one critic suggesting that even a small number of pleiotropic loci can result in false-positive results [29]. Also noted in that article was the worrying observation that this effect appears to increase with larger sample sizes. In a rebuttal to the article, it was stated that horizontal pleiotropy has always been known to impose some limits on MR, but that there are now techniques (examples of which are briefly discussed later in this section) that have been developed to mitigate this problem [30]. A recent analysis came to the conclusion that horizontal pleiotropy was present in approximately 48% of causal links as determined by MR [26]. A concomitant study sought to use
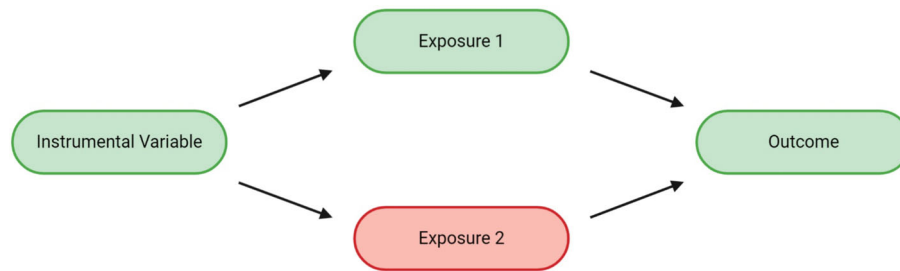
Figure 2. DAG demonstrating horizontal pleiotropy in MR. The IV affects the Outcome through the intended exposure (Exposure 1) but also through another unintended exposure (Exposure 2), thereby violating IV assumption iii (exclusion restriction criterion). Created using BioRender.com.

machine learning to predict an optimal MR method for any specific analysis, in order to improve power and lower false discovery rates; this study found that the preferred method of MR, as predicted by their machine learning framework, involved some degree of horizontal pleiotropy in 90% of tests [16]. From this the conclusion was that horizontal pleiotropy is abundant and unavoidable, and therefore should be evaluated as standard practice when performing MR analysis, that is, choosing a method that is more robust against the effects of horizontal pleiotropy when high levels are detected [16,26].

A number of methods that serve as sensitivity analyses may be used to address the issue of horizontal pleiotropy. One of these involves using multiple SNPs for MR analysis of a single trait and then testing for heterogeneity between these SNPs (since the association between exposure and outcome should be the same for each SNP). Tests such as the between-instrument heterogeneity Q test have been found to work well, especially when the data set is large and both the exposure and outcome data come from the same data set (*i.e.* one-sample MR), but do not indicate the origins of any heterogeneity [31]. Another method that serves as a sensitivity analysis is an adaptation of Egger regression called MR-Egger. It can be used to detect bias that results from horizontal pleiotropy based on the assumption that any pleiotropic effects from IVs on the outcome are independent of the exposure; this is known as the INstrument Strength INdependent of Direct Effect (InSIDE) assumption and is considered to be a weaker version of the exclusion restriction criterion [32]. This assumption allows for pleiotropy in all IVs, but results in less precise analysis of the effects of the exposure results due to reduced statistical power [33].

The Mendelian Randomisation Pleiotropy RESidual Sum and Outlier (MR-PRESSO) test has been developed recently and involves the detection of pleiotropy through comparing the observed distance of all variants to the expected exposure-outcome regression line under the assumption of no horizontal pleiotropy. It uses this to determine outlier variants and then calculates the distortion by comparing causal estimates prior to and following outlier removal [26]. It was found that MR-PRESSO could minimise and correct for horizontal pleiotropy in some cases, but only when the trait responsible for the pleiotropy was known *a priori* [26]. However, outlier detection techniques such as MR-PRESSO also require at least 50% of IVs to be

valid (not horizontally pleiotropic), pleiotropy to be balanced, and the InSIDE assumption to hold [26].

Removing outliers using the median (as opposed to the mean conventionally used in MR-Egger) or mode-based estimate (MBE) methods through the ZEro Modal Pleiotropy Assumption (ZEMPA) of Wald ratio estimates may also mitigate against pleiotropy in MR analyses [22,34]. The weighted median method gives consistent results when at least 50% of the IVs are valid and has been found to be as efficient as the IVW method [22]. Under the ZEMPA, MBE methods can infer a causal effect, even if the majority of IVs are invalid, and can give less biased results than other methods, but generally they have less power to detect causality [34]. However, methods that involve removing outliers should be used carefully, as they are essentially cherry picking the data and potentially excluding SNPs that are biologically relevant [35].

Linkage disequilibrium

Linkage disequilibrium (LD) is defined as the non-random association of alleles at genetic loci that are close to one another on a chromosome. Therefore, violation of the assumptions underlying MR may occur if a SNP being used as an IV is in LD with a SNP that affects the outcome via an independent exposure (Figure 3) [13].

Similarly to horizontal pleiotropy, LD is a common occurrence and should be accounted for in MR analysis through the selection or prioritising of appropriate SNPs [35,36]. As well as setting a maximum pairwise LD threshold for SNP inclusion, methods such as penalised logistic regression have been described as a means of selecting SNPs based on the knowledge of LD [37,38]. A Bayesian test that may be used to determine whether associations are the result of co-localised SNPs could also possibly reduce bias from LD in MR analysis [39]. However, attempts to mitigate LD in cases where it is strong are potentially irrelevant, as it may be impossible to determine which of the IVs is responsible for the effects seen.

Population stratification

Another issue that may arise in MR analysis is population stratification; this is defined as the systematic difference in allele frequencies between subgroups within a population. In the context of genetic association studies such as MR, this difference may serve as a confounding
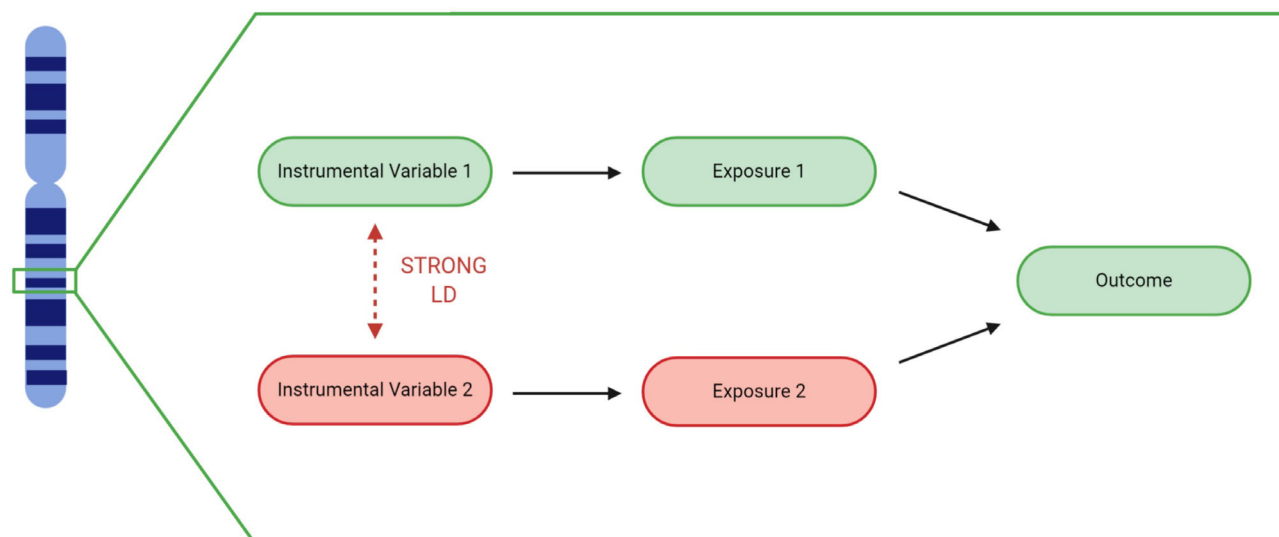
Figure 3. Schematic and DAG demonstrating LD in MR. Instrumental Variables 1 and 2 are located in proximity to one another on a chromosome and are in strong LD (red-dashed double-headed arrow). They both affect the Outcome but are doing so independently through Exposures 1 and 2 respectively, hence violating IV assumption iii. Created using BioRender.com.

variable and lead to spurious associations between SNPs and the outcome [41]. Although some suggest that the bias that could result from population stratification in well-designed epidemiological studies of cancer is relatively small [41], the total effect when combined with bias resulting from other MR limitations may be enough to skew results. One such example of population stratification came from a recent study that found an association between genetic variants and location of birth in UK Biobank genetic data [42].

A number of solutions have been proposed to mitigate the potential bias of population stratification. The seemingly most straightforward solution would be to only use individuals with a relatively homogeneous genetic ancestry. However, another problem frequently seen in MR analysis, low statistical power, is alleviated by using large sample sizes from GWAS data sets. The use of these larger data sets generally results in reduced genetic homogeneity; therefore, it is necessary to find the correct balance between these issues when performing MR.

Another technique that has been suggested involves using principal components to correct for population stratification [43]. Unfortunately, this too results in reduced statistical power. A method using linear mixed models attempts to account for population stratification while maintaining high statistical power through modelling genotype markers together [44]. However, as with horizontal pleiotropy and LD, these methods can only partially address the issue of population stratification [42,45,46].

Trait heterogeneity

The fact that SNPs tend not to affect traits as a whole, but rather certain aspects of traits, poses an issue in MR analysis [14]. It means that each IV may vary in terms of the 'percentage' of a phenotype that it accounts for, or only be causal for a trait in the presence of other relevant SNPs (Figure 4). The purpose of MR is to establish

causality, and so trait heterogeneity can make this very difficult. The only current solution to this problem is to attain a better understanding of the way SNPs affect biological pathways that are linked to the outcome. This may then allow for a specific weighting to be given to the causality of the SNP, as opposed to being able to only infer causality [14]. However, achieving accurate causality weightings based on underlying biology may prove extremely difficult in practice.

Complexity of association

Given that the pathways involved in biological processes are so complex, results that are obtained from large GWAS can often appear to be counterintuitive [14]. This problem extends to MR analysis, and so misunderstanding of the underlying biology may result in incorrect interpretation. One such example of this came from a study investigating the relationship between alcohol consumption and oesophageal cancer risk through examination of genetic isoforms of aldehyde dehydrogenase 2 (*ALDH2*). Individuals who have ALDH2 isoforms that result in an inactive protein are unable to process acetaldehyde, a metabolite of alcohol; they develop symptoms such as dysphoria, nausea, and a flushing reaction in response to alcohol consumption, and therefore tend to consume very low levels [47,48]. MR analysis showed that individuals who were homozygous for the inactive protein had a lower risk of cancer, but also produced the paradoxical finding that heterozygotes had a higher risk than homozygotes for the active form [48,49]. The cause of this erroneous finding was due to varied levels of self-reported alcohol intake, with the results being skewed by heterozygotes who were heavier drinkers than active homozygotes [48]. As with trait heterogeneity, improved understanding of biological pathways and further investigation into paradoxical results are necessary to ensure that the results of MR analysis are accurate.
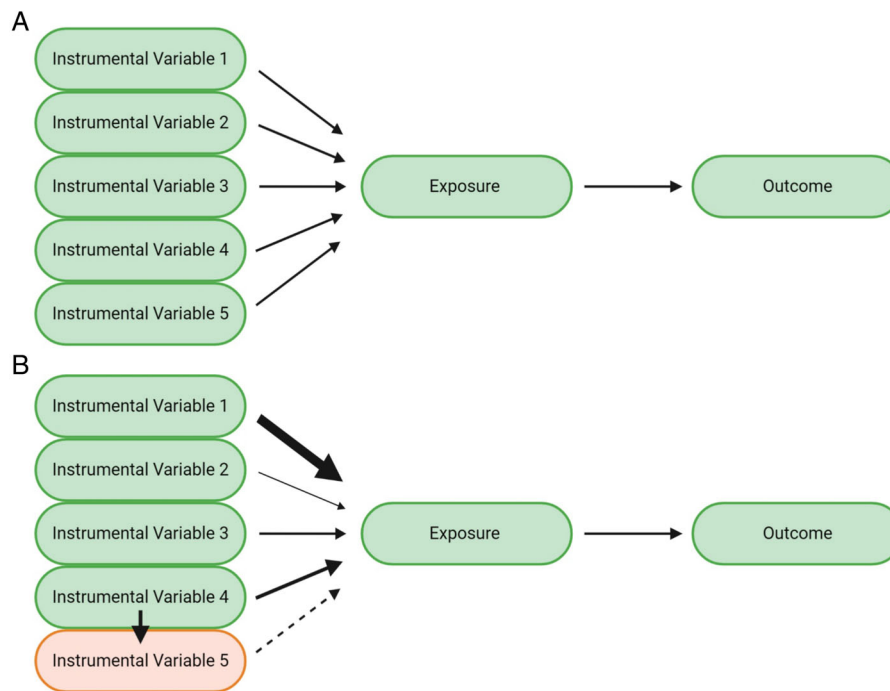
Figure 4. DAGs demonstrating (A) 'simplistic' and (B) 'realistic' conditions encountered in MR due to trait heterogeneity. In 'simplistic' condition (A), Instrumental Variables 1–5 are assumed to have an equal effect on the Exposure and are independent of one another. In 'realistic' condition (B), Instrumental Variables 1–5 have varying effects on the Exposure (as signified by arrow thickness) and their effects on the Exposure may be contingent upon other Instrumental Variables. As an example, in condition B, Instrumental Variable 5 will only act on the Exposure if Instrumental Variable 4 acts on the Exposure (as signified by black arrow linking the variables to one another, orange box and black-dashed arrow from Instrumental Variable 5 to the Exposure). Created using BioRender.com.

Dynastic effects

Dynastic effects may also limit the effectiveness of MR analysis; these effects describe scenarios where the phenotype of the parents has a direct effect on the phenotype of the offspring. In other words, the genotype of the parents affects the outcome of the offspring via means excluding the offspring genotype. This means that the SNPs, exposures, and outcomes of the previous generation may act as confounding variables in the generation of interest [50]. An example of this could be that well-educated parents may not only be able to pass on genes to their offspring that are conducive to higher intelligence, but also create a superior learning environment, for instance, by paying for access to better schools or additional tutoring. Within-family methods using individual-level or summary-data MR may remove some of the bias resulting from dynastic effects [50].

Critical period effects

An issue may arise with MR analysis if the exposure only induces the outcome during a specific period of time during life. The MR analysis will detect this causal link but not be able to distinguish the 'critical period,' unlike RCTs [48]. This means that following MR analysis, any attempts to clinically intervene on an exposure may only be successful if undertaken during this critical period, and the possibility of underestimating the cumulative effects of a lifelong exposure may also occur [48]. A potential solution to this problem may be to use several different epidemiological

approaches to 'triangulate' on the periods of time where presence of the exposure is most likely to lead to the outcome, that is, cancer [48,51]. Another possible way to avoid bias resulting from critical period effects would be to perform a negative control MR analysis; this is performed by measuring the effects of the exposure on the outcome during different hypothetical critical periods to determine whether there is an actual critical period [48].

Weak instrument bias

Weak instrument bias describes a scenario where the IVs appear not to be strongly linked to the exposure and, for example, explain only a small part of the resulting phenotype [13,52]. This then leads to a bias towards the confounded observational association or the null hypothesis, respectively, depending on whether one- or two-sample MR was used [40,48,53]. The direction of bias in one-sample MR becomes evident through examination of the Wald ratio; the IV-outcome coefficient remains constant, but the IV-exposure coefficient is lowered due to weak instruments, meaning that the exposure-outcome coefficient is incorrectly overestimated (Figure 5). Therefore, depending on the type of MR analysis performed, it may be the case that causal risk factors are given a disproportionately high or low weighting. The first stage F-statistic regression of the exposure on the IV is generally used to define strength, with a score lower than 10 defining an instrument as being weak [52,54,55].

One suggested method to reduce weak instrument bias is to increase the sample size, as the F-statistic is dependent
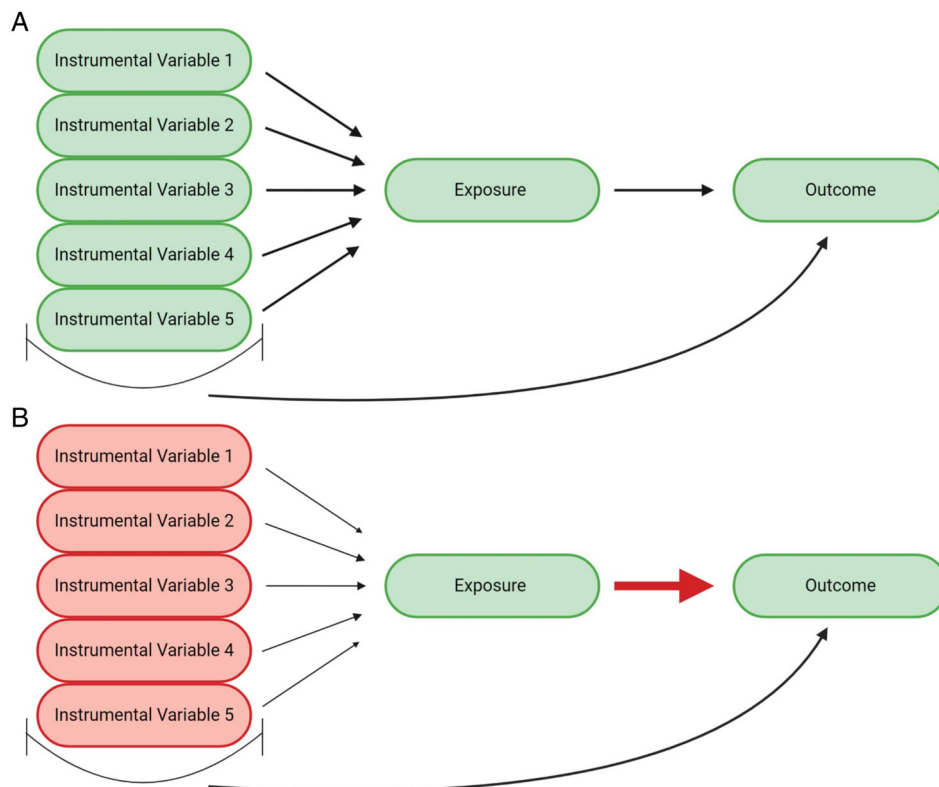
Figure 5. DAGs demonstrating (A) 'ideal' and (B) 'suboptimal' conditions encountered in MR due to weak instrument bias. In 'ideal' condition A, Instrumental Variables 1–5 are strongly associated with the Exposure, resulting in the accurate quantification of the Exposure-Outcome coefficient. In 'suboptimal' condition B, Instrumental Variables 1–5 are weak (signified by red boxes and thin arrows linking them to the Exposure) resulting in an incorrect overestimation of the Exposure-Outcome coefficient (thick red arrow), while the Instrumental Variable-Outcome coefficient is constant in both conditions (curved black arrows). Created using BioRender.com.

on sample size [55]. Another technique based on increasing the F-statistic involves more stringent selection of the IVs used, that is, excluding IVs that explain very little of the resulting phenotype [55]. The 'sample size' solution appears to be easier to implement, as, for example, genetic variants may explain as little as 1% of a phenotype; this makes it difficult to be stringent with the number of IVs used, while ensuring that as much of the phenotype as possible is accounted for [14]. However, reliance on increasing the F-statistic alone does not appear to be useful in terms of reducing bias, as approaches where studies with F-values lower than 10 have been excluded from meta-analysis had no effect [55]. The same study concluded that using tests to measure the strength of IVs does not give a true indication of the variable strength; attempts to omit 'weak' variables on the basis of their F-statistic values were deemed to be overly simplistic and may actually result in more bias, due to the varying magnitude of effects depending on whether the observed F-statistic was greater or less than the expected F-value [55].

As mentioned previously, using fewer IVs may minimise bias from weak instruments. This appears to be due to a higher chance of imbalance of the confounding variables between subgroups defined by the IVs [55]. This applies even to IVs that are biologically relevant; while inclusion of these variables increases precision, it may also increase bias [55].

Another potential mechanism of reducing weak instrument bias involves adjusting the MR analysis to include measured covariates; this has been found to increase precision, especially when covariates explain outcome variation, as well as lower weak instrument bias [55]. The inclusion of covariates has been found to increase the F-statistic and reduce median bias from IV estimators, even for stronger IVs. Similarly, use of allelic scores, which are single variables that encompass multiple IVs associated with a risk factor, provides a way to use fewer instruments. This can reduce weak instrument bias while ensuring that a larger proportion of the resulting phenotype is accounted for [56].

Gaining a better understanding of the magnitude of effect that a genotype has on a phenotype by utilising information from different studies may also reduce weak instrument bias in meta-analyses. It has been shown that combining sub-studies within a meta-analysis, as opposed to combining summary estimates, gives pooled estimates with reduced bias. In addition, the assumption of common genetic effects across studies appears to be able to eradicate weak instrument bias [55].

Winner's curse

The term 'winner's curse' originated from the field of economics and describes a scenario where an individual

who places the winning bid at an auction will tend to be 'cursed', having been forced to overvalue the item in order to win, and will therefore make a net loss [57,58]. In the context of GWAS, winner's curse refers to an analogous occurrence that it is often only the lead SNP with the smallest *P*-value that is reported, whereas other significant SNPs may not even be mentioned [14]. For one-sample MR analysis, this may result in an overestimation of the lead SNP-exposure effect due to chance correlation between instrumental and confounding variables during the discovery stage of the GWAS [40,48].

The ratio of coefficients or Wald method describes a means of determining the causal effect of an exposure on an outcome and is calculated by dividing the IV-outcome coefficient by the IV-exposure coefficient [20,21]. Therefore, if the IV was discovered in data set separate from the MR analysis data set, the presence of winner's curse will result in an inflated IV-exposure coefficient and a reduced Wald ratio, that is, an underestimated causal effect [14,48]. In the instance where the GWAS discovery and MR analysis data sets are the same, both the IV-outcome coefficient and the IV-exposure coefficient will be overestimated, also potentially resulting in an incorrect Wald ratio, albeit most likely a less incorrect one [14].

One suggested method to alleviate the issue of winner's curse is to perform two-sample MR analysis [48,59]. Overestimation resulting from winner's curse will tend to mean that any effects from confounding variables are being underestimated. In cases where the outcome is binary (*e.g.* disease causal versus non-causal), bias may be avoided if only control participants are used in the discovery data set, but if cases are used in addition to controls, this will result in weak instrument bias [60]. Determining the bias in MR analysis resulting from overlapping data sets is difficult and so exercising caution before including IVs that lie close to the significance threshold is recommended, as well as avoiding data-driven approaches for acquiring IVs [60].

Low statistical power

Another inherent dilemma in MR analysis is low statistical power; this is because the IVs or SNPs used will usually explain only a fraction of the phenotype [14]. Estimates for causality are also imprecise, which results in larger confidence intervals, and makes determining a causal effect through MR analysis more difficult [40].

The solutions to low statistical power in MR analysis mirror those used to resolve weak instrument bias. One such method involves increasing the sample sizes using large GWAS consortia and summary data sets [14,40,48,61,62]. However, a study in 2014 concluded that smaller sample sizes may not necessarily prevent MR analyses from attaining sufficient power [62]. Another approach involves combining individual IVs into an allelic score that serves as a single

IV; performing this allows for greater coverage of the phenotype [14,56,63].

Collider/selection bias

In the context of MR analysis, a collider is a variable that is causally downstream of both the exposure and the outcome. Collider bias can occur when statistical adjustments, or conditioning on the collider, are attempted [40,64,65]. This means that sample selection may introduce bias into MR analysis. For example, if a collider influences participation in a study, then it is possible to overestimate a spurious causal link between the IV and the outcome [64]. One situation in which this could arise is in the study of cancer progression; it is important to take into consideration the selection bias that will occur given that having the cancer in question will be a prerequisite for entry into the MR study.

Collider bias may act either towards the association or towards the null depending on the IV; if the variable is involved in cancer incidence but not progression, only focussing on cases will lead to an overestimation [66]. Conversely, only studying cases for an IV responsible for both cancer incidence and progression results in collider bias towards the null [66].

Selection bias is considered to be a form of collider bias and there are a number of scenarios whereby it is likely to occur in MR analysis. One such example is performing MR in the context of disease progression, where in order to be included in a study, participants must have been diagnosed previously with the disease in question. In this situation, if the exposure is a risk factor for the outcome (*i.e.* the disease), then participation in the study is being affected by a collider and will therefore result in bias [65]. Other forms of collider bias that may occur in cancer MR analyses include survivor bias, which may lead to an overestimation of the effect in the general population based on observations made in the elderly, and subpopulation bias, which may occur following the recruitment of hospitalised patients [65].

A number of methods have been suggested to mitigate collider/selection bias. One such method involves using any known parameters such as disease prevalence to estimate the bias and using analytical formulae or inverse probability weighting to correct for it [66–68]. Inverse probability weighting involves taking into account underrepresented cases in a data set and gives them more weight in the analysis, making the assumption that these cases are likely to be more prevalent in the general population [65,67]. To prevent extremely rare cases from being granted a very large weight in the analysis, weights may be trimmed to a threshold *post hoc*, although it is recommended that the initial generation of weights is performed accurately in order to avoid trimming [65,69]. In simulations, inverse probability weighting reduced selection bias when the model was correctly specified and where there was a large selection effect. However, it induced worse bias than the initial selection bias when the selection effect was small and the effects of trimming were only prevalent in extreme

cases [65]. The authors concluded that moderate selection bias tends not to affect MR estimates too severely relative to other biases, and that using inverse probability weighting to rectify bias can be effective but only if the magnitude of the bias is known beforehand [65].

Another suggested method is to check for associations between genetic variants and the outcome in disease progression; any associations found between the same variants and disease incidence should be noted, as they may imply potential collider bias [66]. Associations between genetic variants and confounding variables that are found within the chosen MR analysis data set(s) but not in the general population may also suggest that both are causally upstream of the disease and that there may be collider bias [66].

Practical implementation of MR

There are several tools available for conducting MR analysis. Packages in R such as MendelianRandomization and TwoSampleMR may be used to perform two-sample MR using GWAS-derived summary-level data [70,71]. MendelianRandomization allows for the implementation of methods such as MR-Egger and weighted median, and provides a graphical output of causal estimates for each method used [70].

MR-Base is a web application that integrates a database of GWAS results with R packages (such as TwoSampleMR) to automate two-sample MR and allows for the performance of the entire MR workflow [71]. First, appropriate IVs may be obtained from exposure GWAS, and then the effects of these IVs on the outcome may be assessed; the next step involves harmonising the aforementioned data (ensuring that the effect allele of the SNP is the same for both the exposure and the outcome) prior to the performance of MR analysis [71]. Post-MR, sensitivity analyses such as funnel plots and leave-one-out analysis may be implemented to identify limitations [71]. Funnel plots may be used to visualise the relationship between the IV strength and the causal estimate, thereby highlighting directional pleiotropy in MR-Egger sensitivity analyses [32]. Leave-one-out analysis involves repeating the MR analysis with a different individual IV removed each time, allowing for the identification of outliers that are potentially skewing the data [40].

MR analyses with multiple IVs are typically performed using the IVW method. Alternative robust methods of MR may be used to detect pleiotropy and therefore serve as a form of quality control (QC) for the initial analysis [17]. For example, if MR-Egger analysis results in an effect estimate that differs greatly from the original method chosen, it may indicate that many of the IVs are not valid and the originally obtained causal effect estimate is not robust [32,72]. An MR-Egger intercept that is far from 0 suggests that the IV-exposure and IV-outcome relationships are not linear, and that there may be directional pleiotropy and hence no genuine

causal effect [72]. Equally for median- and mode-based methods, which are more robust against outliers than MR-Egger, a causal estimate that is similar to the IVW and MR-Egger is indicative of an accurate estimation of causality [17,22,34]. Using MR-Egger, a median-based approach and a mode-based approach have been recommended as QC, given that they each require different assumptions to hold, but differences in the estimates from these analyses do not necessarily imply a lack of causality [17]. In addition, a test for heterogeneity among MR estimates such as the between-instrument Q test is recommended [17,31]. Although some heterogeneity may be expected even if all IVs are valid, a large degree of heterogeneity (Q test P-value <0.05) may result in a less reliable causal effect estimate [31]. This is particularly relevant when there are strong outliers, which may represent pleiotropic IVs, or if the causal effect is dependent on a small number of IVs [17]. The I^2 statistic may be used to check for weak instrument bias in MR-Egger analysis; values closer to 1 suggest that bias is not present, while values closer to 0 may be indicative of weak instrument bias [72,73]. For some of the limitations of MR analysis, QC exists and may be used to identify and attempt to correct for biases. However, other problems (*e.g.* complexity of association) are difficult to detect, and the unique nature of MR analyses means that there may not be a singular right answer with regard to the way QC is performed.

MR studies in cancer have now been performed using exposures such as alcohol consumption, vitamin D levels, and body mass index (BMI), among many others (Figure 6) [74,75]. MR has been used to conclude that higher levels of alcohol consumption increase risk in esophageal and head and neck cancer [49,76]. However, for colorectal cancer, MR analyses have been conflicted on whether there is a causal association [77,78]. MR studies attempting to establish causal links between vitamin D levels and cancer have also given mixed results. There is evidence from MR both for and against a vitamin D causal association for ovarian and prostate cancer [79–83], and no association was found for breast, colorectal, lung, neuroblastoma, and pancreatic cancers in MR studies [80,81,83–86]. Differing results between cancer types have also been found in MR analyses attempting to relate BMI to cancer risk. Higher BMI has been associated with increased colorectal, endometrial, gastric, kidney, and ovarian cancer risk [87–95], but has been found to decrease breast cancer risk [89,96–98]. MR studies using increased BMI as an exposure have found evidence both for and against a causal association with lung cancer [89,92,99,100], as well as no causal association with prostate cancer [89,101].

These studies demonstrate the complexity involved in using MR to study cancer. Alongside some notably consistent studies (Figure 6), many exposures have been found to only be causal in some types of cancer, and studies in a single cancer type have not consistently shown a causal link, or even a consistent direction of causality. In addition, some exposures have been
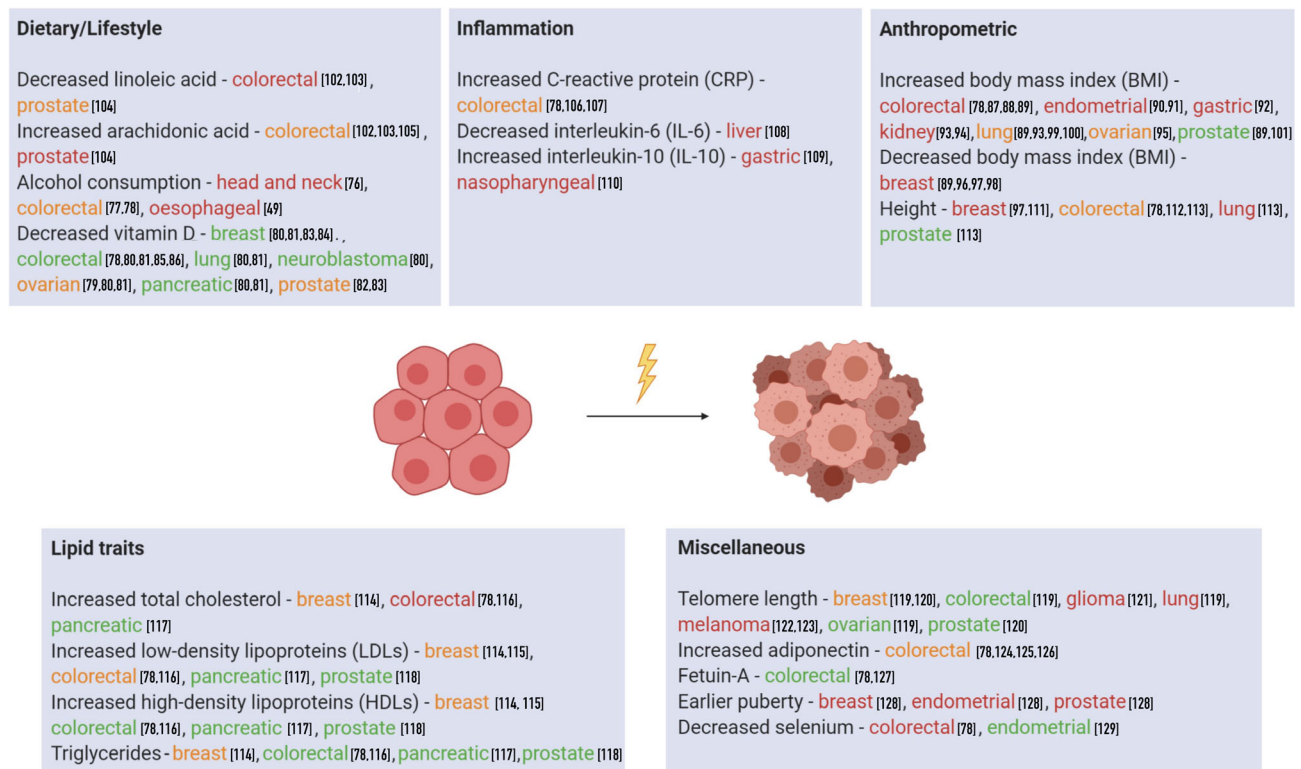
Figure 6. Exemplar results from MR studies to test for causal relationships between different phenotypic traits (exposures) and cancers (outcomes). For cancers in red, MR results were consistent with causality of the phenotypic trait, whereas cancers in green were not deemed to be caused by the phenotypic trait. Cancers in orange had conflicting results, that is, disagreement between MR studies on whether there was a causal link, disagreement between MR studies on the direction of causality, or phenotypic trait examined in MR studies was found to only cause (1) cancer in certain subpopulations (*e.g.* women/elderly patients), (2) specific cancer subtypes (*e.g.* ER-positive breast cancers), or (3) cancer only when analysis was conducted using certain data sets. Created using BioRender.com.

deemed to cause cancer in only certain subpopulations (*e.g.* women or elderly patients), only specific cancer subtypes, or only when MR is conducted using certain data sets. Whether these results reflect differences in underlying biology is often unclear.

Conclusions

MR is an established epidemiological tool used to infer causality that can provide both ethical and practical advantages over alternatives such as RCTs. MR has recently gained popularity alongside the rise in GWAS for normal and disease traits. In the study of cancer, especially, using traditional RCTs to potentially expose a case group to an entity that may increase the likelihood of developing disease is not viable. MR also mitigates the issues of reverse causation and confounding variables, as germline variants are randomly distributed according to Mendel's laws of heritability. Tools have been developed that allow many of the basic MR analyses to be performed efficiently by non-expert biostatisticians. However, despite these advantages over other epidemiological techniques, MR has many limitations of its own.

Many of the previously discussed limitations are due to violation of the assumptions that underlie MR and a recurring issue involves the accurate determination of IVs.

Valid IVs are important for accurate MR analysis and failure to use them may result in missed or spurious associations. Issues such as trait heterogeneity and population stratification may also have major effects on MR, even if these problems are not unique to MR. The suggestion that they can be minimised with a greater understanding of the biological pathways involved is something easier said than done. Two-sample MR appears to be a more practical approach than one-sample MR, as it helps somewhat to address limitations such as low statistical power, winner's curse, and weak instrument bias; one straightforward way that two-sample MR accomplishes this is that it facilitates the acquisition of greater amounts of data. Combining IVs to give allelic scores is useful for the avoidance of weak instrument bias and increases the probability that more of the phenotype of interest is being accounted for. A number of sensitivity analyses and inverse probability weighting may be used to attempt to correct for bias post-MR. Unfortunately, these techniques appear to only be truly effective if the magnitude and direction of the bias are known *a priori*. Studies using MR simulations are a useful tool that will help those performing MR understand where limitations such as selection/collider bias are more likely to occur, but it is important to ensure that the simulations reflect actual MR analysis as much as possible.

In summary, it appears that the limitations of MR analysis currently mean that it must be used very

cautiously to determine causality in cancer, and this is reflected in the conflicting results observed in many cancer MR studies. Many of the exposures used thus far in cancer MR studies that are suggested to be causal following analysis are challenging, if not impossible, to verify using clinical trials. Orthogonal evidence of causality is required but often not available, and it is difficult to verify whether MR studies have been conducted with enough power to infer causality. Confirming the causality of exposures that lend themselves to trials is arguably a priority, and this could be used to infer the reliability of MR studies generally. MR analysis has the potential to be an effective tool in cancer research when it is combined with other epidemiological techniques and follow-up biological work.

Acknowledgements

Author contributions statement

HG wrote and edited the manuscript. IT designed the review outline and edited the manuscript. Both authors approved the final version.

References

1. Peto R, Doll R, Buckley JD, *et al*. Can dietary beta-carotene materially reduce human cancer rates? *Nature* 1981; **290:** 201–208.
2. Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* 1994; **330:** 1029–1035.
3. Hennekens CH, Buring JE, Manson JE, *et al*. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med* 1996; **334:** 1145–1149.
4. Omenn GS, Goodman GE, Thornquist MD, *et al*. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N Engl J Med* 1996; **334:** 1150–1155.
5. Tatsioni A, Bonitsis NG, Ioannidis JP. Persistence of contradicted claims in the literature. *JAMA* 2007; **298:** 2517–2526.
6. Sheehan NA, Didelez V, Burton PR, *et al*. Mendelian randomisation and causal interference in observational epidemiology. *PLoS One* 2008; **5:** e177.
7. Cornish AJ, Tomlinson IPM, Houlston RS. Mendelian randomisation: a powerful and inexpensive method for identifying and excluding non-genetic risk factors for colorectal cancer. *Mol Aspects Med* 2019; **69:** 41–47.
8. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986; **1:** 507–508.
9. Trompet S, Jukema JW, Katan MB, *et al*. Apolipoprotein e genotype, plasma cholesterol, and cancer: a Mendelian randomization study. *Am J Epidemiol* 2009; **170:** 1415–1421.
10. Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. *Bone Marrow Transplant* 1991; **7:** 9–12.
11. Burnett AK, Wheatley K, Goldstone AH, *et al*. The value of allogeneic bone marrow transplant in patients with acute myeloid leukaemia at differing risk of relapse: results of the UKMRC AML 10 trial. *Br J Haematol* 2002; **118:** 385–400.
12. Keavney B. Commentary: Katan's remarkable foresight: genes and causality 18 years on. *Int J Epidemiol* 2004; **33:** 11–14.
13. Lawlor DA, Harbord RM, Sterne JA, *et al*. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; **27:** 1133–1163.
14. Haycock PC, Burgess S, Wade KH, *et al*. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr* 2016; **103:** 965–978.
15. Hartwig FP, Davies NM, Hemani G, *et al*. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol* 2016; **45:** 1717–1726.
16. Hemani G, Bowden J, Haycock P, *et al*. Automating Mendelian randomization through machine learning to construct a putative casual map of the human phenome. *bioRxiv* 2017: 173682. https://doi.org/10.1101/173682. Not peer reviewed.
17. Burgess S, Davey Smith G, Davies NM, *et al*. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* 2019; **4:** 186.
18. Teumer A. Common methods for performing Mendelian randomization. *Front Cardiovasc Med* 2018; **5:** 51.
19. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; **37:** 658–665.
20. Wald A. The fitting of straight lines if both variables are subject to error. *Ann Math Stat* 1940; **11:** 284–300.
21. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stats Methods Med Res* 2017; **26:** 2333–2355.
22. Bowden J, Davey Smith G, Haycock PC, *et al*. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; **40:** 304–314.
23. Szumillas M. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry* 2010; **19:** 227–229.
24. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015; **181:** 251–260.
25. Zuber V, Colijn JM, Klaver C, *et al*. Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nat Commun* 2020; **11:** 29.
26. Verbanck M, Chen CY, Neale B, *et al*. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018; **50:** 693–698.
27. Sivakumaran S, Agakov F, Theodoratou E, *et al*. Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet* 2011; **89:** 607–618.
28. Gratten J, Visscher PM. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med* 2016; **8:** 78.
29. Pickrell JK. Fulfilling the promise of Mendelian randomization. *bioRxiv* 2015: 018150. https://doi.org/10.1101/018150. Not peer reviewed.
30. Davey SG. Mendelian randomization: a premature burial? *bioRxiv* 2015: 021386. https://doi.org/10.1101/021386. Not peer reviewed.

31. Greco MFD, Minelli C, Sheehan NA, *et al*. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat Med* 2015; **34:** 2926–2940.

32. 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:** 512–525.

33. Davies NM, Holmes MV, Davey SG. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018; **362:** k601.

34. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; **46:** 1985–1998.

35. Hemani G, Bowden J, Davey SG. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018; **27:** R195–R208.

36. Purcell S, Neale B, Todd-Brown K, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81:** 559–575.

37. Verzilli C, Shah T, Casas JP, *et al*. Bayesian meta-analysis of genetic association studies with different sets of markers. *Am J Hum Genet* 2008; **82:** 859–872.

38. Ayers KL, Cordell HJ. SNP selection in genome-wide and candidate gene studies via penalized logistic regression. *Genet Epidemiol* 2010; **34:** 879–891.

39. Giambartolomei C, Vukcevic D, Schadt EE, *et al*. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; **10:** e1004383.

40. Zheng J, Baird D, Borges MC, *et al*. Recent developments in Mendelian randomization studies. *Curr Epidemiol Rep* 2017; **4:** 330–345.

41. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 2000; **92:** 1151–1158.

42. Haworth S, Mitchell R, Corbin L, *et al*. Apparent latent structure within the UKBiobank sample has implications for epidemiological analysis. *Nat Commun* 2019; **10:** 333.

43. Price AL, Patterson NJ, Plenge RM, *et al*. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38:** 904–909.

44. Loh PR, Tucker G, Bulik-Sullivan BK, *et al*. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015; **47:** 284–290.

45. Abdellaoui A, Hugh-Jones D, Yengo L, *et al*. Genetic correlates of social stratification in Great Britain. *Nat Hum Behav* 2019; **3:** 1332–1342.

46. Koellinger PD, de Vlaming R. Mendelian randomization: the challenge of unobserved environmental confounds. *Int J Epidemiol* 2019; **48:** 665–671.

47. Peng GS, Yin SJ. Effect of the allelic variants of aldehyde dehydrogenase ALDH2*2 and alcohol dehydrogenase ADH1B*2 on blood acetaldehyde concentrations. *Hum Genomics* 2009; **3:** 121–127.

48. Yarmolinsky J, Wade KH, Richmond RC, *et al*. Causal inference in cancer epidemiology: what is the role of Mendelian randomization? *Cancer Epidemiol Biomarkers Prev* 2018; **27:** 995–1010.

49. Lewis SJ, Davey SG. Alcohol, ALDH2, and esophageal cancer: a meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. *Cancer Epidemiol Biomarkers Prev* 2005; **14:** 1967–1971.

50. Davies NM, Howe LJ, Brumpton B, *et al*. Within family Mendelian randomization studies. *Hum Mol Genet* 2019; **28:** R170–R179.

51. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016; **45:** 1866–1886.

52. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* 2011; **30:** 1312–1323.

53. Bound J, Jaeger D, Baker R. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc* 1995; **90:** 443–450.

54. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 2002; **20:** 518–529.

55. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011; **40:** 755–764.

56. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013; **42:** 1134–1144.

57. Capen EC, Clapp RV, Campbell WM. Competitive bidding in high-risk situations. *J Petrol Technol* 1971; **23:** 641–653.

58. Thaler R. Anomalies: the winner's curse. *J Econ Perspect* 1988; **2:** 191–202.

59. Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol* 2016; **45:** 908–915.

60. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016; **40:** 597–608.

61. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013; **42:** 1497–1501.

62. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol* 2014; **43:** 922–929.

63. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011; **40:** 740–752.

64. Munafò MR, Tilling K, Taylor AE, *et al*. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2018; **47:** 226–235.

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

66. Paternoster L, Tilling K, Davey SG. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: conceptual and methodological challenges. *PLoS Genet* 2017; **13:** e1006944.

67. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res* 2013; **22:** 278–295.

68. Yaghootkar H, Bancks MP, Jones SE, *et al*. Quantifying the extent to which index event biases influence large genetic association studies. *Hum Mol Genet* 2017; **26:** 1018–1030.

69. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. *PLoS One* 2011; **6:** e18174.

70. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017; **46:** 1734–1739.

71. Hemani G, Zheng J, Elsworth B, *et al*. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018; **7:** e34408.

72. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-egger method. *Eur J Epidemiol* 2017; **32:** 377–389.

73. Bowden J, Del Greco MF, Minelli C, *et al*. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-egger regression: the role of the I2 statistic. *Int J Epidemiol* 2016; **45:** 1961–1974.

74. Pierce BL, Kraft P, Zhang C. Mendelian randomization studies of cancer risk: a literature review. *Curr Epidemiol Rep* 2018; **5:** 184–196.

75. Benn M, Nordestgaard BG. From genome-wide association studies to Mendelian randomization: novel opportunities for understanding cardiovascular disease causality, pathogenesis, prevention, and treatment. *Cardiovasc Res* 2018; **114:** 1192–1208.

76. Boccia S, Hashibe M, Gallì P, *et al*. Aldehyde dehydrogenase 2 and head and neck cancer: a meta-analysis implementing a Mendelian randomization approach. *Cancer Epidemiol Biomarkers Prev* 2009; **18:** 248–254.

77. Wang J, Wang H, Chen Y, *et al*. Alcohol ingestion and colorectal neoplasia: a meta-analysis based on a Mendelian randomization approach. *Colorectal Dis* 2011; **13:** e71–e78.

78. Cornish AJ, Law PJ, Timofeeva M, *et al*. Modifiable pathways for colorectal cancer: a Mendelian randomisation analysis. *Lancet Gastroenterol Hepatol* 2020; **5:** 55–62.

79. Ong JS, Cuellar-Partida G, Lu Y, *et al*. Association of vitamin D levels and risk of ovarian cancer: a Mendelian randomization study. *Int J Epidemiol* 2016; **45:** 1619–1630.

80. Dimitrakopoulou VI, Tsilidis KK, Haycock PC, *et al*. Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study. *BMJ* 2017; **359:** j4761.

81. Ong JS, Gharahkhani P, An J, *et al*. Vitamin D and overall cancer risk and cancer mortality: a Mendelian randomization study. *Hum Mol Genet* 2018; **27:** 4315–4322.

82. Bonilla C, Gilbert R, Kemp JP, *et al*. Using genetic proxies for life-course sun exposure to assess the causal relationship of sun exposure with circulating vitamin D and prostate cancer risk. *Cancer Epidemiol Biomarkers Prev* 2013; **22:** 597–606.

83. Jiang X, Dimou NL, Al-Dabhani K, *et al*. Circulating vitamin D concentrations and risk of breast and prostate cancer: a Mendelian randomization study. *Int J Epidemiol* 2019; **48:** 1416–1424.

84. Wang S, Huo D, Kupfer S, *et al*. Genetic variation in the vitamin D related pathway and breast cancer risk in women of African ancestry in the root consortium. *Int J Cancer* 2018; **142:** 36–43.

85. Theodoratou E, Palmer T, Zgaga L, *et al*. Instrumental variable estimation of the causal effect of plasma 25-hydroxy-vitamin D on colorectal cancer risk: a Mendelian randomization analysis. *PLoS One* 2012; **7:** e37662.

86. He Y, Timofeeva M, Farrington SM, *et al*. Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. *BMC Med* 2018; **16:** 142.

87. Thrift AP, Gong J, Peters U, *et al*. Mendelian randomization study of body mass index and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 2015; **24:** 1024–1031.

88. Jarvis D, Mitchell JS, Law PJ, *et al*. Mendelian randomisation analysis strongly implicates adiposity with risk of developing colorectal cancer. *Br J Cancer* 2016; **115:** 266–272.

89. Gao C, Patel CJ, Michailidou K, *et al*. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. *Int J Epidemiol* 2016; **45:** 896–908.

90. Nead KT, Sharp SJ, Thompson DJ, *et al*. Evidence of a causal association between insulinemia and endometrial cancer: a Mendelian randomization analysis. *J Natl Cancer Inst* 2015; **107:** djv178.

91. Painter JN, O'Mara TA, Marguart L, *et al*. Genetic risk score Mendelian randomization shows that obesity measured as body mass index, but not waist:hip ratio, is causal for endometrial cancer. *Cancer Epidemiol Biomarkers Prev* 2016; **25:** 1503–1510.

92. Mao Y, Yan C, Lu Q, *et al*. Genetically predicted high body mass index is associated with increased gastric cancer risk. *Eur J Hum Genet* 2017; **25:** 1061–1066.

93. Brennan P, McKay J, Moore L, *et al*. Obesity and cancer: Mendelian randomization approach utilizing the FTO genotype. *Int J Epidemiol* 2009; **38:** 971–975.

94. Johansson M, Carreras-Torres R, Scelo G, *et al*. The influence of obesity-related factors in the etiology of renal cell carcinoma-a Mendelian randomization study. *PLoS Med* 2019; **16:** e1002724.

95. Dixon SC, Nagle CM, Thrift AP, *et al*. Adult body mass index and risk of ovarian cancer by subtype: a Mendelian randomization study. *Int J Epidemiol* 2016; **45:** 884–895.

96. Guo Y, Warren Andersen S, Shu XO, *et al*. Genetically predicted body mass index and breast cancer risk: Mendelian randomization analyses of data from 145,000 women of European descent. *PLoS Med* 2016; **13:** e1002105.

97. Qian F, Wang S, Mitchell J, *et al*. Height and body mass index as modifiers of breast cancer risk in BRCA1/2 mutation carriers: a Mendelian randomization study. *J Natl Cancer Inst* 2019; **111:** 350–364.

98. Shu X, Wu L, Khankari NK, *et al*. Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis. *Int J Epidemiol* 2019; **48:** 795–806.

99. Carreras-Torres R, Haycock PC, Relton CL, *et al*. The causal relevance of body mass index in different histological types of lung cancer: a Mendelian randomization study. *Sci Rep* 2016; **6:** 31121.

100. Carreras-Torres R, Johansson M, Haycock PC, *et al*. Obesity, metabolic factors and risk of different histological types of lung cancer: a Mendelian randomization study. *PLoS One* 2017; **12:** e0177875.

101. Davies NM, Gaunt TR, Lewis SJ, *et al*. The effects of height and BMI on prostate cancer incidence and mortality: a Mendelian randomization study in 20,848 cases and 20,214 controls from the PRACTICAL consortium. *Cancer Causes Control* 2015; **26:** 1603–1616.

102. May-Wilson S, Sud A, Law PJ, *et al*. Pro-inflammatory fatty acid profile and colorectal cancer risk: a Mendelian randomisation analysis. *Eur J Cancer* 2017; **84:** 228–238.

103. Khankari NK, Banbury BL, Borges MC, *et al*. Mendelian randomization of circulating polyunsaturated fatty acids and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 2020; **29:** 860–870.

104. Khankari NK, Murff HJ, Zeng C, *et al*. Polyunsaturated fatty acids and prostate cancer risk: a Mendelian randomisation analysis from the PRACTICAL consortium. *Br J Cancer* 2016; **115:** 624–631.

105. Isom CA, Shrubsole MJ, Cai Q, *et al*. Arachidonic acid and colorectal adenoma risk: a Mendelian randomization study. *Clin Epidemiol* 2018; **11:** 17–22.

106. Nimptsch K, Aleksandrova K, Boeing H, *et al*. Association of CRP genetic variants with blood concentrations of C-reactive protein and colorectal cancer risk. *Int J Cancer* 2015; **136:** 1181–1192.

107. Wang X, Dai JY, Albanes D, *et al*. Mendelian randomization analysis of C-reactive protein on colorectal cancer risk. *Int J Epidemiol* 2019; **48:** 767–780.

108. Tian G, Mi J, Wei X, *et al*. Circulating interleukin-6 and cancer: a meta-analysis using Mendelian randomization. *Sci Rep* 2015; **5:** 11394.

109. Niu W, Pang Q, Lin T, *et al*. A causal role of genetically elevated circulating interleukin-10 in the development of digestive cancers: evidence from Mendelian randomization analysis based on 29,307 subjects. *Medicine* 2016; **95:** e2799.

110. Qu K, Pang Q, Lin T, *et al*. Circulating interleukin-10 levels and human papilloma virus and Epstein-Barr virus-associated cancers: evidence from a Mendelian randomization meta-analysis based on 11,170 subjects. *Onco Targets Ther* 2016; **9:** 1251–1267.

111. Zhang B, Shu XO, Delahanty RJ, *et al*. Height and breast cancer risk: evidence from prospective studies and Mendelian randomization. *J Natl Cancer Inst* 2015; **107:** djv219.

112. Thrift AP, Gong J, Peters U, *et al*. Mendelian randomization study of height and risk of colorectal cancer. *Int J Epidemiol* 2015; **44:** 662–672.

113. Khankari NK, Shu XO, Wen W, *et al*. Association between adult height and risk of colorectal, lung, and prostate cancer: results from meta-analyses of prospective studies and Mendelian randomization analyses. *PLoS Med* 2016; **13:** 1002118.

114. Beeghly-Fadiel A, Khankari NK, Delahanty RJ, *et al*. A Mendelian randomization analysis of circulating lipid traits and breast cancer risk. *Int J Epidemiol* 2019. https://doi.org/10.1093/ije/dyz242.

115. Nowak C, Ärnlöv J. A Mendelian randomization study of the effects of blood lipids on breast cancer risk. *Nat Commun* 2018; **9:** 3957.

116. Rodriguez-Broadbent H, Law PJ, Sud A, *et al*. Mendelian randomisation implicates hyperlipidaemia as a risk factor for colorectal cancer. *Int J Cancer* 2017; **140:** 2701–2708.

117. Carreras-Torres R, Johansson M, Gaborieau V, *et al*. The role of obesity, type 2 diabetes, and metabolic factors in pancreatic cancer: a Mendelian randomization study. *J Natl Cancer Inst* 2017; **109:** djx012.

118. Bull CJ, Bonilla C, Holly JM, *et al*. Blood lipids and prostate cancer: a Mendelian randomization analysis. *Cancer Med* 2016; **5:** 1125–1136.

119. Zhang C, Doherty JA, Burgess S, *et al*. Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. *Hum Mol Genet* 2015; **24:** 5356–5366.

120. Luu HN, Long J, Wen W, *et al*. Association between genetic risk score for telomere length and risk of breast cancer. *Cancer Causes Control* 2016; **27:** 1219–1228.

121. Walsh KM, Codd V, Rice T, *et al*. Longer genotypically-estimated leukocyte telomere length is associated with increased adult glioma risk. *Oncotarget* 2015; **6:** 42468–42477.

122. Iles MM, Bishop DT, Taylor JC, *et al*. The effect on melanoma risk of genes previously associated with telomere length. *J Natl Cancer Inst* 2014; **106:** dju267.

123. Rachakonda S, Kong H, Srinivas N, *et al*. Telomere length, telomerase reverse transcriptase promoter mutations, and melanoma risk. *Genes Chromosomes Cancer* 2018; **57:** 564–572.

124. Pei Y, Xu Y, Niu W. Causal relevance of circulating adiponectin with cancer: a meta-analysis implementing Mendelian randomization. *Tumour Biol* 2015; **36:** 585–594.

125. Song M, Gong J, Giovannucci EL, *et al*. Genetic variants of adiponectin and risk of colorectal cancer. *Int J Cancer* 2015; **137:** 154–164.

126. Nimptsch K, Song M, Aleksandrova K, *et al*. Genetic variation in the ADIPOQ gene, adiponectin concentrations and risk of colorectal cancer: a Mendelian randomization analysis using data from three large cohort studies. *Eur J Epidemiol* 2017; **32:** 419–430.

127. Nimptsch K, Aleksandrova K, Boeing H, *et al*. Plasma fetuin-A concentration, genetic variation in the AHSG gene and risk of colorectal cancer. *Int J Cancer* 2015; **137:** 911–920.

128. Day FR, Thompson DJ, Helgason H, *et al*. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* 2017; **49:** 834–841.

129. Kho PF, Glubb DM, Thompson DJ, *et al*. Assessing the role of selenium in endometrial cancer risk: a Mendelian randomization study. *Front Oncol* 2019; **9:** 182.