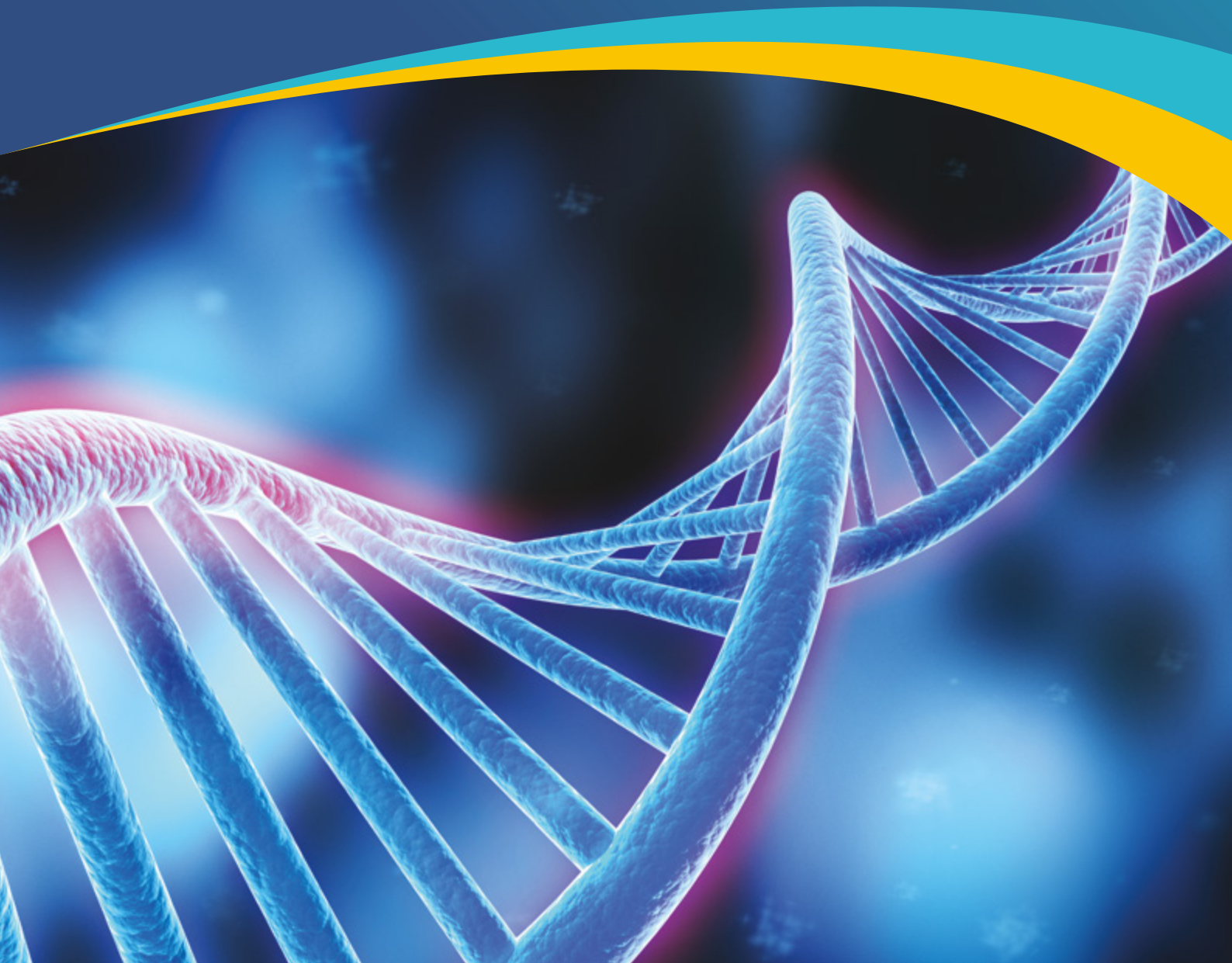# Making sense of Mendelian randomisation and its use in health research

A short overview

Sean Harrison, Laura Howe and Alisha R. Davies

Research and Evaluation Division
Knowledge Directorate
Public Health Wales NHS Trust
Number 2 Capital Quarter
Tyndall Street
Cardiff CF10 4BZ

Tel: +44 (0)29 2022 7744

Email: phw.research@wales.nhs.uk

🐦 @PHREWales

f @PublicHealthWales

www.bristol.ac.uk/integrative-epidemiology/mr-methods

Contents

1. Introduction

Population health research seeks to better understand the causes and consequences of ill health. It generates evidence to help inform decision making and action to protect good health, and ensures the capacity to benefit is equitable across all individuals irrespective of circumstances.

The most common studies which explore the relationship between causes (i.e. exposures) and outcomes are observational in design. They examine the characteristics of individuals, their experiences and surroundings using routinely available information (e.g. health and non-health data) or data collected for the specific question (e.g. in surveys). But this approach faces two key challenges (*reverse causality and confounding* – both described in more detail below) which may limit our ability to identify whether there is a true association between exposure and outcome. Mendelian randomisation is a different statistical approach harnessing the information contained within our genome to help overcome these challenges.

In this overview, we explain the principles of Mendelian randomisation, and illustrate how it is being used to challenge and inform population health.

This short report is intended for a non-specialist audience (including but not limited to health economists, academics in other areas of expertise, public health practitioners or policy makers) to increase awareness of the approach, its strengths and weaknesses and its application.

This is not a full review of Mendelian randomisation, nor is it a technical reference document, but we have provided references for further reading for those who are interested in learning more.

2. Why do we need to understand cause and effect?

Understanding the relationships between cause (i.e. exposure) and effect (i.e. outcome) is essential to informing policy and practice.

Reliable evidence is needed to help understand what we need to focus efforts on, if intervening could lead to benefits, and the potential magnitude of the benefits of intervention.

For example, an observational study may find that obesity is associated with poor educational outcomes. However, this conclusion may not be correct if there are other factors associated with both exposure and outcome. For example, people living in less affluent areas have lower educational attainment, and are also at higher risk of obesity (**Figure 1a**). The higher risk of obesity amongst those with lower education attainment is likely to be due, in part, to deprivation (the challenge termed 'confounding') (**Figure 1b**). A further complication is that association may be in the other direction i.e. poor educational outcome is the cause of obesity, due to low self-esteem and increased uptake of health harming behaviours leading to obesity (the challenge termed 'reverse causality'), (**Figure 1c**).

Figure 1. Confounding and reverse causality in the association between obesity and education outcomes

Figure 1a

Question: does obesity cause poor educational outcomes, does poor education cause obesity, or does something else cause both?

Figure 1b

Confounding: deprivation causes both obesity and poor educational outcomes, making it seem like obesity causes poor educational outcomes.

Figure 1c

Reverse causality: poor education causes obesity, rather than obesity causing poor educational outcomes.

2.1. What is confounding?

Confounding variables are factors that we know about, we can measure and control for in statistical analysis. However, even when we know about a confounder, we are unlikely to have measured it perfectly, especially for complex things such as socioeconomic circumstances. There will also be confounders we do not know about, have not measured and have not considered. This means there is still some confounding (residual confounding) in most observational studies.

2.2. What is reverse causality?

Reverse causality is assigning a direction to an association between two variables incorrectly, for example by stating that lung cancer causes cigarette smoking, rather than the reverse. Causality is very difficult to determine for many associations, even when confounding is not an issue. In the example above (Figure 1), did obesity lead to poor educational outcomes, or did poor educational outcomes lead to obesity? Statistical tests can, in general, only tell you whether two variables are associated – to determine causality, you need to consider study design, biological plausibility, and other relevant information.

2.3. Approaches to overcome confounding and reverse causality, and the limitations

The gold standard method to address both confounding and causality is a randomised controlled trial (RCT). RCTs usually work by randomly splitting a sample of participants into two or more groups, giving one group the exposure (usually an intervention), and the other group(s) (control group) a different exposure (or intervention). An intervention can be a drug, a diet, information; anything that might change the individual's outcome in some way. Because which group a participant ends up in is random, the only difference between the two groups should be whether they received the intervention or not, which means there should be no residual confounding. In addition, there can be no reverse causality, as the intervention clearly comes before the outcome of interest.

However, for many research questions, it is impossible or unethical to randomly assign the exposure. For example, it would not be possible nor acceptable to randomly allocate obesity.

3. What is Mendelian randomisation and how might it help?

Mendelian randomisation is a statistical approach that uses genetics to provide information about the relationship between an exposure and outcome (a type of instrumental variable analysis) (1).

The combination of *genetic variant*s a person receives from their parents is randomly assigned at conception - making Mendelian randomisation *a natural experiment* which is potentially less likely than observational studies to be subject to confounding and reverse causality, and relatively quicker and easier than RCT designs to complete a large scale research study.

Mendelian randomisation as an approach has become widespread in the past decade. The expansion of large scale *genome wide association studies* (GWAS) has developed our knowledge about the genetic determinants of human characteristics and health conditions, and large scale research studies that include genetic data, e.g. UK Biobank (2) [**www.ukbiobank.ac.uk**] which have made data more available.

3.1. What is a genetic variant?

Genetic variants are small parts of the genome (**Box 1**) which can be closely related to human characteristics (e.g. height, weight, blood pressure) and health conditions (e.g. diabetes, coronary heart disease, asthma). There can be multiple genetic variants for a single characteristic or health condition; for example, over 900 variants are known to affect our body mass index (BMI) (3). Collectively these genetic variants explain only a small amount of the variation in BMI across a population (6% in total), they are not deterministic (i.e. all people with genetic variants that predispose to high BMI do not *necessarily* have a high BMI), and their existence does not negate the important environmental influences on BMI. Nonetheless they can be used to help understand the consequences of BMI, since people who have differing genetic variants related to BMI will, on average, have different BMIs but should not differ with respect to other characteristics (See **Box 2**).

Box 1. Genome: A genome is an organism's complete set of DNA, including all of its genes.


Each genome contains all of the information needed to build and maintain that organism. In humans, a copy of the entire genome – more than 3 billion DNA base pairs – is contained in all cells that have a nucleus.

Box 2. Mendelian randomisation and body mass index (BMI)

If people with a higher number of genetic variants that predispose to higher BMI also have a higher rate of hypertension, this is evidence that higher BMI causes the risk of hypertension to increase. This inference is possible because people with higher and lower genetic propensity for high BMI should not differ according to background, family, socioeconomic or other factors, i.e. there is no confounding, and hypertension cannot affect genetic propensity for high BMI, i.e. there is no reverse causality. In 2017, a study showed that the odds of having hypertension increased 1.64 times per ~5 kg/m^2 increase in BMI (32).

This is illustrated in **Figure 2**.

Figure 2. Illustration of using Mendelian randomisation to show BMI increases systolic blood pressure
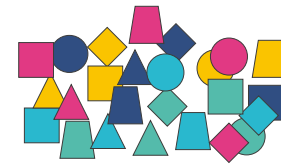
General Population

1. People in the general population have many differences that could confound the relationship between BMI and hypertension (shown here as different shapes).

Very low   Low   Medium   High   Very high
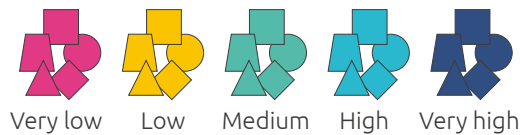
Genetic propensity for high BMI

2. However, these differences should not affect their genetic propensity for having a high or low BMI (shown here as different colours).
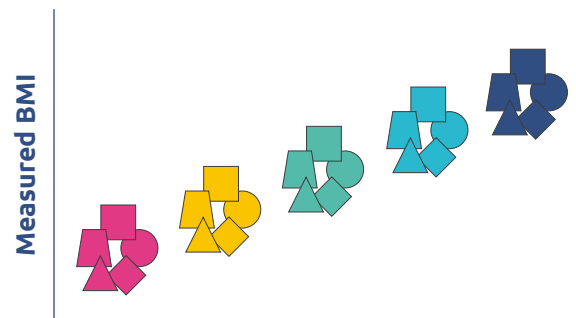
Measured BMI

Genetic propensity for high BMI

3. On average, people with a higher genetic propensity towards a high BMI will have higher measured BMI values

Measured rate of hypertension
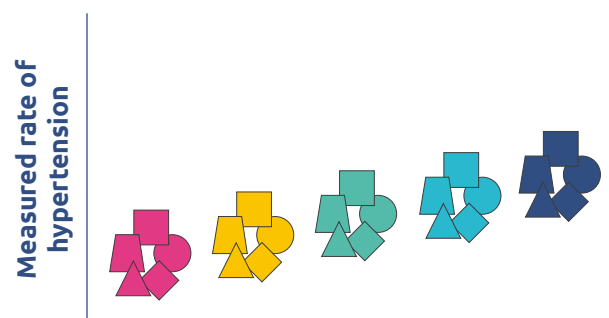
Genetic propensity for high BMI

4. If, on average, people with higher genetic propensities towards a high BMI also have a higher rate of hypertension, then we have evidence that BMI causally increases hypertension risk, as genetic propensity can't be affected by confounders or hypertension itself.

3.2. When can Mendelian randomisation be applied?

Mendelian randomisation is necessarily limited to exposures we know have a genetic basis – we can't, for instance, use Mendelian randomisation to assess whether living in cities causes cancer since there are no known genetic variants likely to directly predispose towards living in a city. Although we have known genetic variants for many health-related factors and conditions, many things of great interest to public health are still difficult to study with Mendelian randomisation; for example, little is known about the genetic determinants of falls in the elderly or back pain. There are genetic variant markers for over 2,000 characteristics and health conditions currently available (more information available from the NHGRI-EBI GWAS Catalog (4)).

3.3. What are the strengths of Mendelian randomisation compared to traditional study designs?

Whilst observational studies can adjust for confounding it can only do that for variables which have been measured (see Section 1.1). The advantage of Mendelian randomisation is that it may be able to overcome the challenge of confounding between the exposure and outcome, regardless of whether the confounders were measured, because those confounders should not influence a person's genetic predisposition towards the exposure. However, there remain some important potential sources of cofounding with the approach (see **Section 3.4**).

Mendelian randomisation is likely to overcome the challenge of reverse casualty. Any genetic variant may cause the outcome, which in turn causes the exposure, rather than vice-versa as expected. However, this risk can be lessened if a mechanism of action of the genetic variant can be established. For example, a prominent genetic variant for smoking heaviness affects a nicotinic receptor, making reverse causality much less likely (5).

Additionally, bi-directional Mendelian randomisation can be used to assess in which direction causality is most likely to flow (1). In bi-directional Mendelian randomisation, Mendelian randomisation analyses are performed in both directions (exposure to outcome, and outcome to exposure). If the exposure truly causes the outcome, then the genetic variants associated with the exposure will cause both the exposure and the outcome, but the genetic variants associated with the outcome will only cause the outcome, not the exposure. Conversely, if the outcome causes the exposure, the reverse will be true. Bi-directional Mendelian randomisation can thus only be performed if there are genetic variants for both the exposure and outcome.

For example, a 2019 study using bi-directional Mendelian randomisation looking at BMI and polycystic ovary syndrome found the genetic variants for BMI were associated with polycystic ovary syndrome, but the genetic variants for polycystic ovary syndrome were not associated with BMI, implying a higher BMI causes polycystic ovary syndrome and not vice-versa (6). Bi-directional Mendelian randomisation, however, does not work with exposures which have not been reported to associate with specific genetic variants, for example living in cities or back pain.

## 3.4. What are the limitations of Mendelian randomisation?

Given the small effects of genetic variants, **large sample sizes are needed** to determine an effect. This has been made easier by the existence of large studies such as UK Biobank (2), but the **populations recruited within such large studies are often not representative of the wider population** (7,8). This can introduce selection bias into the results. The approach to overcome this in the future will be to encourage diversity in population cohorts, and the development of statistical approaches to account for potential selection bias in the analysis.

Like all statistical methods, Mendelian randomisation has certain **assumptions** which must be considered for it to be applied appropriately.

These are;

- **The genetic variants must be associated with the exposure, (1) in Figure 3**
This assumption can and should be verified by testing the association between the genetic variants and the exposure within the data being used.

- **The genetic variants must not be directly associated with the outcome, (2) in Figure 3**
We can use biological knowledge about the genetic variants to tell us something about how likely this is, known as horizontal pleiotropy. (There are also a range of sensitivity analyses that can detect and adjust for pleiotropy (9–14)).

- **The genetic variant must not be associated with any potential confounder (3) in Figure 3**
Confounders can be associated with genetic variants if the choice of partner is non-random, for example if people were more likely to have children with people with similar BMI levels to themselves, or from similar populations (population stratification) – see below.

Figure 3. The assumptions of Mendelian randomisation
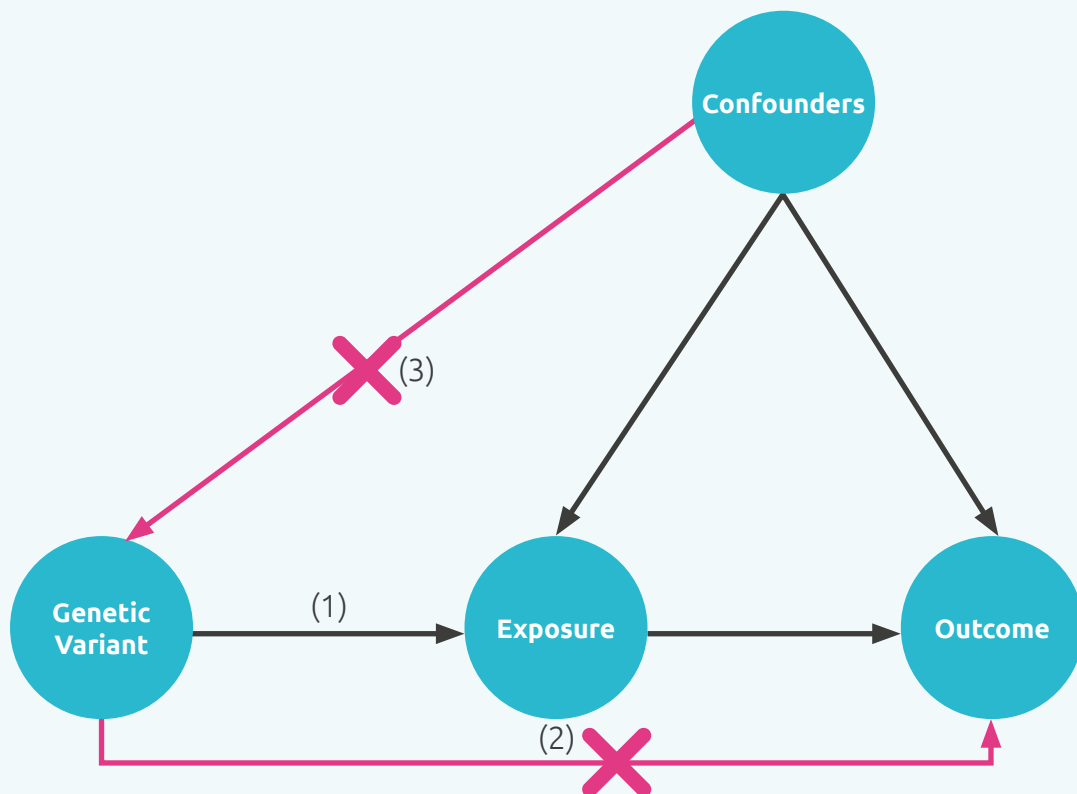
More recently there is increasing recognition that the presumed random distribution of genetic variants (see **Section 2.1**) may not always hold as family and social factors can influence the genes we inherit, for example:

- **People do not choose partners at random**, but rather select those who are more alike than would be expected (termed assortative mating). For example, people who drink alcohol choose partners who also drink alcohol (11), or people select partners who have a similar environmental/educational background to themselves (15).

- **Parental (non-genetic) characteristics can influence outcomes**. For example, more educated parents will support their own child's education (termed dynastic effects). One way to address this bias is to compare outcomes across siblings (within-sibling analysis) (9,16), assuming siblings grow up in the same family-environment (see **Box 3**).

- There are **geographical patterns in genetics** which may make it difficult to identify true differences in results (17) and methods to account for this are not readily available.

Box 3. Example of a within-sibling analysis

An MR study in 2019 showed that higher BMI reduced educational attainment across the population in the UK (33). However, this association disappeared when a within-sibling analysis was conducted, which accounted for the non-random distribution of genetic variants (16). This means that the parental influence on a child's education is likely the driver behind the observed association between BMI and educational attainment, not a direct causal link between BMI and education.

However, because within-sibling or within-family MR studies necessarily require siblings or families, these analyses have less power to detect effects than standard MR analyses that have fewer restrictions.

3.5. Key considerations when interpreting the results

Perhaps the most important assumption for Mendelian randomisation is that we assume the effect of changing the genetic variants is the same as the effect of changing the modifiable exposure we are interested in through other means. For example, we assume that changing BMI by changing the number of BMI-increasing alleles a person has is going to have the same effects on the outcome as changing BMI through interventions we are interested in from a public health perspective, such as diet or exercise.

For BMI, we know that genetic variants influence BMI through a range of mechanisms, including dietary intake, energy expenditure, and energy storage. This assumption, sometimes called 'gene-environment equivalence', is therefore likely to be reasonable. Thinking about this assumption requires biological knowledge of the ways in which genetic variants affect the modifiable exposure.

The more biologically proximal the exposure is (how close the exposure is to a protein made by only a few genes), the more likely the gene-environment equivalence assumption is to hold.

When we are studying the effects of individual proteins or metabolites, which are generally coded by a well-defined and well-understood set of genetic variants, this assumption is usually straightforward and non-controversial. For more complex exposures, such as educational attainment (for which we now know about over 1,271 genetic variants that together explain 11-13% of the variation in educational levels (18)), thinking about gene-environment equivalence is more complicated. However, it's worth noting that even for a complex multi-factorial exposure like educational attainment, results of Mendelian randomisation studies looking at the causal effects of education on health and other outcomes have generally been consistent with other studies using different natural experiments, such as using raising of the school leaving age (19).

Mendelian randomisation can give us an estimate of the causal effect of a change in an exposure of interest. However, this effect estimate does not necessarily map directly on to the potential effect of a clinical or public health intervention. An important reason for this is that Mendelian randomisation estimates a 'lifetime' effect of the exposure. For example, if our exposure is systolic blood pressure and our outcome is cardiovascular disease events, we can estimate the effect of having a lower systolic blood pressure by 10 mmHg across the entire life course. Whilst this can tell us something about whether reducing systolic blood pressure is likely to have an effect on the frequency of cardiovascular disease events, it cannot give us an exact answer about how much we will change the incidence of cardiovascular disease events if we administer antihypertensives to adults in mid-life.

Likewise, if we use Mendelian randomisation to assess the causal impact of BMI on income, we can say something about whether interventions that reduce population levels of BMI are likely to have an effect on income, but we cannot say anything about how feasible those interventions are, and how big an effect on income any specific intervention targeted at a specific age group is likely to have.

One way of viewing this is that Mendelian randomisation tells us something about 'states', but not about specific interventions that alter those states.

As such Mendelian randomisation is another tool to help us understand the causal links between exposures and outcomes, and that understanding can help to direct where action is needed to effect change in an outcome.

To help support those who are new to Mendelian randomisation the following boxes include key points to interpret Mendelian randomisation within population health research including;

- use of Mendelian randomisation (Box 4)

- why observational and Mendelian randomisation studies might give different answers (Box 5).

- factors to consider when reading and interpreting Mendelian randomisation studies (Box 6)

Box 4. Benefits and limitations of the application of Mendelian randomisation

Strengths

Confounding

- The chance of confounding from both imperfectly measured and unknown confounders may be reduced for Mendelian randomisation studies, compared with other observational study designs

Reverse causality

- The chance of reverse causality is equally reduced (for most associations), exposures are unlikely to cause differences in DNA.

A balance between causality and data availability

- While well-conducted RCTs have no risk of confounding or reverse causality, they are expensive, time-consuming, limited in what data can be acquired, and not always ethical or practical

- Mendelian randomisation studies strike a balance between RCTs, which are causal but limited, and observational studies, which are not causal but able to collect far more information. Mendelian randomisation studies can use observational data where the participants genomes are mapped (i.e genotyped), and the exposures have genetic variants, to improve estimate causal relationships

Limitations

Low statistical power

- Mendelian randomisation studies typically have much lower statistical power than other observational study designs of a similar sample size, and so generally require more participants

- This is being addressed by collecting data from huge numbers of participants, for example UK Biobank.

Reverse causality

- Although typically much less of an issue than in observational studies, reverse causality can still exist in Mendelian randomisation studies when the genetic variants we think act only on the exposure, actually act on the outcome to cause the exposure

- This can be addressed by assessing the biological mechanisms of the genetic variants, and, if possible, using bi-directional Mendelian randomisation

Confounding

- Again, confounding of the exposure-outcome relationship can be controlled for in Mendelian randomisation studies, but can suffer from genotype-outcome confounding due to assortative mating and population stratification

- This can be addressed using within-siblings and within-family Mendelian randomisation study designs

Horizontal pleiotropy

- The genetic variants associated with the exposure may affect the outcome through other variables than the exposure

- This can be addressed by assessing the biological mechanisms of the genetic variants, as well as a range of additional analyses that have been designed specifically to test for horizontal pleiotropy.

Interpretation

- Since genetic variants are present from conception, Mendelian randomisation estimates are for lifetime exposure, which may not map well onto specific interventions to change the exposure.

Box 5. Potential reasons why observational and Mendelian randomisation studies might give different answers.

The observational results may be confounded, or there may be reverse causation, while the Mendelian randomisation results are unconfounded and causal.

The populations may differ, which means both the observational and Mendelian randomisation results are true, but for different populations, and may not generalize beyond those populations.

The Mendelian randomisation results may be less precise than the observational results, even if more participants are included, meaning there is more variability around the estimated effect in the Mendelian randomisation result.

The Mendelian randomisation result may not satisfy the Mendelian randomisation assumptions, and so the genetic variant(s) may not be good enough, there may be horizontal pleiotropy, or there may be confounding from family effects or population stratification.

Box 6. Things to consider when reading and interpreting an Mendelian randomisation study, adapted from *Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians* by Davies et al. 2018 (20)

Core assumptions

- Are the genetic variants robustly associated with the exposure?
- Are the genetic variants associated with potential confounders of the exposure-outcome association, and do the authors present this information?
- Do the authors present alternative Mendelian randomisation analyses to assess the risk of horizontal pleiotropy, such as MR Egger, median, and mode estimators, or use of "negative control" populations?

Methods reporting

- Were the GWAS and analytical samples drawn from the same population?

Data presentation

- Do the authors compare estimates from genetic and conventional observational analyses?
- Do the authors attempt to remove genetic variants that show signs of pleiotropy?
- Do the authors provide the data that they used to allow researchers to reproduce their findings?

Limitations

- Have all the relevant limitations (see **Box 3**) been acknowledged, and could these limitations meaningfully affect the results?

Clinical implications

- Do the results triangulate with other forms of evidence, e.g. RCTs, observational studies?

4. Where has Mendelian randomisation influenced the discussion or provided answers?

Mendelian randomisation has now been applied to hundreds of different research questions. Below are three examples of key findings.

4.1. Alcohol and Cardiovascular Disease

Alcohol consumption is difficult to measure accurately – for different reasons, people may either misremember or misreport how much they have consumed over time. Associations between alcohol and health outcomes are also likely to be heavily confounded, since lots of variables affect how much people drink, and these factors also plausibly affect many health outcomes of interest. People who abstain from alcohol completely are often very different from those who drink moderate quantities. When the outcome is a health condition, there may also be reverse causality – illness, even preclinical phases of many health conditions, can lead people to reduce their alcohol consumption. As such, the traditional J-shape curve you see when looking at observational data between alcohol and health conditions (especially cardiovascular disease, see **Figure 4**) may be due to both confounding and reverse causality, rather than any beneficial aspects of moderate alcohol consumption.

Although RCTs have been conducted, randomising people to alcohol or no alcohol, they tend to be small studies with a short followup time, and thus are not suitable for examining the effects of alcohol on disease events.

Larger scale and duration RCTs are not feasible; it is very difficult to either force people to drink or not to drink over time, and expensive to follow people for many years. Therefore, Mendelian randomisation can provide insights that RCTs can't.

The alcohol dehydrogenase enzyme is responsible for breaking down alcohol and has genetic variants that affect how well it functions. People with a non-perfect alcohol dehydrogenase enzyme tend to get a "flush" response to alcohol and may get headaches and nausea quickly after drinking alcohol. As such, people with a genetic variation in this gene tend to drink much less throughout their lives than people with fully functioning enzymes.

Thus, the genetic variants that affect the levels of alcohol dehydrogenase in the body can be used in Mendelian randomisation, to understand the effect of differing levels of alcohol consumption on health outcomes.

Holmes and colleagues conducted a Mendelian randomisation study looking at the association between alcohol and cardiovascular disease, and concluded the following (21):

Individuals with a genetic variant associated with non-drinking and lower alcohol consumption had a more favourable cardiovascular profile and a reduced risk of coronary heart disease than those without the genetic variant. This suggests that reduction of alcohol consumption, even for light to moderate drinkers, is beneficial for cardiovascular health.

Thus, this evidence suggests the J-shaped relationship between alcohol and cardiovascular disease seen in observational studies is likely to be driven by confounding and/or reverse causality. Because the Mendelian randomisation analysis is not subject to confounding or reverse causality, and we are reasonably certain that

the genetic variant acts only through alcohol to affect cardiovascular health, we can be more confident in the conclusion than an observational study could be.

Figure 4. J-shaped relationship between alcohol consumption (X-axis) and the relative risk (RR, Y-axis) of coronary heart disease seen in observational data, from (22) [source: https://content.iospress.com/articles/nutrition-and-aging/nua0052#ref041
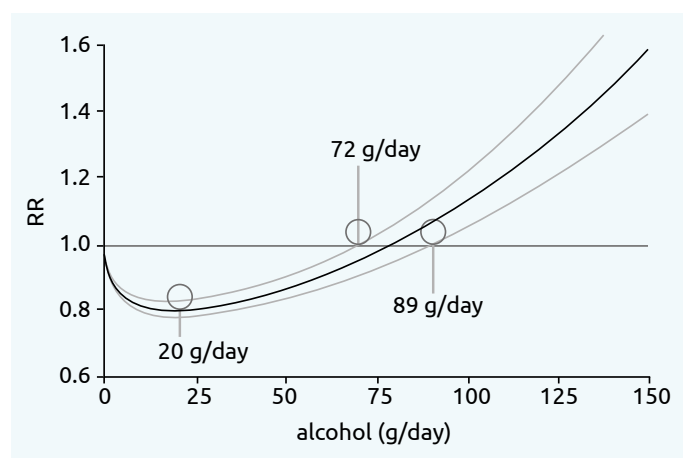
4.2. C-reactive Protein and Coronary Heart Disease

C-reactive protein (CRP) is a protein involved in inflammation that has been used for decades as a marker for cardiovascular risk – the higher a person's CRP, the higher their risk of experiencing cardiovascular disease. Because of this, in the early and mid-2000s, CRP was considered to be a potential drug target, with drugs that lower CRP hypothesised to reduce the risk of developing cardiovascular disease.

Given the only evidence for this possibility came from observational studies, there was a chance that either confounding or reverse causality caused the association between CRP and cardiovascular risk. Mendelian randomisation offered the opportunity to assess this, to evaluate whether the lengthy and costly process of developing and testing a drug to target CRP would be likely to yield benefits.

A Mendelian randomisation study looked at a genetic variant strongly associated with CRP in relation to coronary heart disease risk (23). Their analysis showed that there was no evidence that changes in CRP led to changes in coronary heart disease risk. This was entirely different to the strong positive association between CRP and coronary heart disease seen in observational studies. The team concluded:

We found no association of a genetic variant, which is known to be related to CRP levels, (rs1130864) and having CHD. These findings do not support a causal association between circulating CRP and CHD risk, but very large, extended, genetic association studies would be required to rule this out.

This and similar examples have been used by the pharmaceutical industry to identify and prioritise potential drug targets, diverting resources away from avenues that are not causal and therefore not likely to be fruitful avenues for drug development research.

4.3. Body-mass index and Hospital Costs

Body-mass index (BMI) is another variable that is difficult to study due to both confounding and the difficulty in performing RCTs that solely change BMI – eating healthier or exercising could have independent effects, limiting the ability to pin down the causal association between BMI and any outcome.

However, there are several genetic variants that predispose towards higher or lower BMIs. As such, a recent study has shown that as BMI increases, so do healthcare costs (more than in a comparable observational analysis) (24). This could be used as further evidence for the benefits of reducing BMI across the population.

5. Conclusion

Mendelian randomisation is a versatile tool that can be of great value in many research areas, notably when RCTs are not practical or ethical. Now we are well into the era of large-scale genetic studies, we know of genetic variants that affect many human characteristic and health conditions, and can quickly and easily use this information to answer questions about whether the associations between an exposure and an outcome are causal. Mendelian randomisation, like all statistical methods, requires certain assumptions, and it is important to assess these assumptions as far as possible within the data.
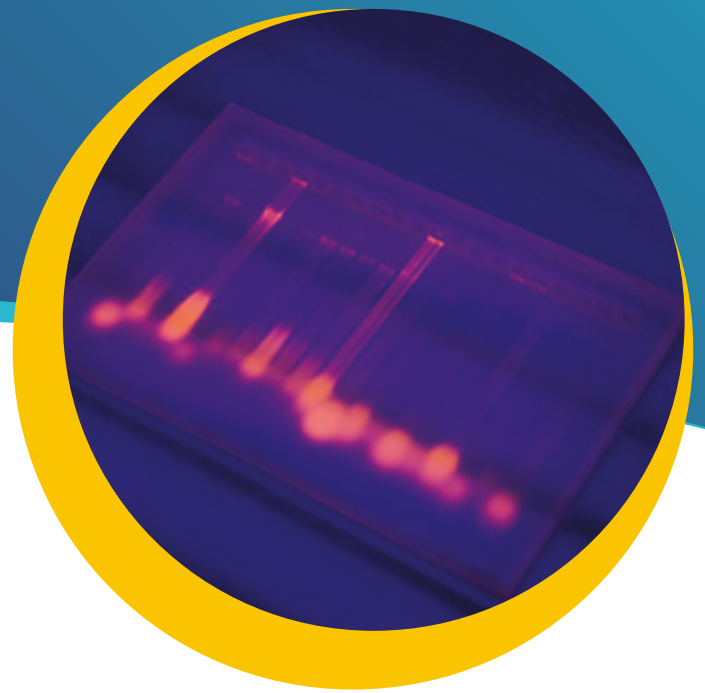
The results from Mendelian randomisation studies represent a lifetime exposure and tell us about 'states' rather than specific interventions, but as Mendelian randomisation can be robust to exposure-outcome confounding and reverse causality, it can answer questions not possible to answer using observational studies. In effect, Mendelian randomisation strikes a balance between RCTs, which are causal but limited, and observational research, which is far less limited but rarely causal. The results of Mendelian randomisation studies (indeed, any study) should be considered within the context of the wider evidence – indeed, triangulation of evidence, where an association is assessed across a range of methods (e.g. Mendelian randomisation, cohort studies, RCTs etc.) can give the most reliable answers to scientific and clinical questions (25).

We hope this guide will inform non-specialist audiences to help raise awareness of the Mendelian randomisation method, its strengths and weaknesses and help others to interpret the findings within the context of wider evidence to inform evidence-based practice and policy.

6. A dictionary of terms for Mendelian randomisation: what does it all mean?

Term	User friendly description	Reference
Assortative mating	Similar people (behaviours or backgrounds) have children together	(9)
Collider bias, also called **selection bias**	Bias in an analysis from adjusting for (conditioning on, or selecting for) a variable that is caused by both the exposure and the outcome. For example, fever is caused by both influenza and food poisoning. If you just look at people with fever, there will be a negative correlation between influenza and food poisoning, since if someone doesn't have influenza, they must have food poisoning (and vice versa), making it look like influenza is protective for food poisoning. This is within people with a fever – if you looked at everyone, there would be no association between influenza and food poisoning.	(26)
Confounding	Bias found when a third variable causes both the exposure and outcome in an analysis.	(27)
Dynastic effects	When parental genotypes directly affect their children's phenotypes, not through genes. e.g. more educated parents support their child's education, or if parental smoking habits influence their child's smoking habits, also known as genetic nurture, or gene-environment correlation.	(16)
Horizontal pleiotropy & pleiotropic effects	When a genetic variant that associates with the exposure affects the outcome through a variable other than the exposure.	(20)
Genetic variant, also called **single-nucleotide polymorphism (SNP)**	A point-change in the genome (DNA) of a person, that may predispose that person to have a certain characteristic or health condition - GWAS typically look at millions of genetic variants to find those that are associated with an exposure.	(28)
Genome wide association studies (GWAS)	A study conducted to find genetic variants associated with an exposure – these studies inform which genetic variants are used in a Mendelian randomisation study. The GWAS and Mendelian randomisation study must be conducted in separate populations to avoid bias (independence of GWAS and Mendelian randomisation datasets).	(29)
Mendelian randomisation, often abbreviated to **MR**	An analysis using genetic variants to estimate the causal effect of an exposure on an outcome	(1)
MR Egger	A sensitivity analysis used to detect horizontal pleiotropy	30)

Term	User friendly description	Reference
Polygenic risk scores (PRS), also called **genetic risk scores (GRS)**	Instead of using individual genetic variants for an exposure, a polygenic risk score is a weighted sum of all genetic variants, condensing all genetic information into a single variable, and improving statistical power.	(31)
Reverse causality	When the outcome causes the exposure, rather than vice-versa as previously supposed.	(1)
Split-sample	The same population can be used to conduct a GWAS and Mendelian randomisation study simultaneously, but only if the population is split in two – one for the GWAS, and one for the Mendelian randomisation analysis, to preserve independence of the GWAS and Mendelian randomisation datasets.	(33)
Within-sibling/ within-family analysis	Estimating the effects of an exposure using genetic variation occurring between siblings or parent-offspring trios.	(16)

7. References

1. Davey Smith G, Hemani G. Mendelian randomisation: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet [Internet]. 2014;23(R1):R89-98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25064373%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4170722

2. Allen NE, Sudlow C, Peakman T, Collins R. UK biobank data: Come and get it. Science Translational Medicine. 2014;6(224).

3. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. Hum Mol Genet. 2018;

4. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42(D1).

5. Gage SH, Jones HJ, Taylor AE, Burgess S, Zammit S, Munafò MR. Investigating causality in associations between smoking initiation and schizophrenia using Mendelian randomisation. Sci Rep. 2017;7.

6. Brower MA, Hai Y, Jones MR, Guo X, Chen YDI, Rotter JI, et al. Bidirectional Mendelian randomisation to explore the causal relationships between body mass index and polycystic ovary syndrome. Hum Reprod. 2019;34(1):127–36.

7. Hughes RA, Davies NM, Smith GD, Tilling K. Selection bias in instrumental variable analyses. bioRxiv [Internet]. 2017;192237. Available from: https://www.biorxiv.org/content/early/2017/09/22/192237

8. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. Am J Epidemiol. 2017;186(9):1026–34.

9. Hartwig FP, Davies NM, Davey Smith G. Bias in Mendelian randomisation due to assortative mating. Genet Epidemiol. 2018;42(7):608–20.

10. 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. 2018;

11. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomisation via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985–98.

12. Bowden J, Smith GD, Burgess S. Mendelian randomisation with invalid instruments: Effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25.

13. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomisation with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304–14.

14. Burgess S, Thompson SG. Multivariable Mendelian randomisation: The use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251–60.

15. Bowden J, Fabiola Del Greco M, Minelli C, Smith GD, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample mendelian randomisation analyses using MR-Egger regression: The role of the I 2 statistic. Int J Epidemiol. 2016;45(6):1961–74.

16. Brumpton B, Sanderson E, Hartwig FP, Harrison S, Vie GÅ, Cho Y, et al. Within-family studies for Mendelian randomisation: avoiding dynastic, assortative mating, and population stratification biases. bioRxiv [Internet]. 2019;602516. Available from: https://www.biorxiv.org/content/10.1101/602516v1?rss=1&utm_source=dlvr.it&utm_medium=twitter

17. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications. bioRxiv [Internet]. 2018;294876. Available from: https://www.biorxiv.org/content/early/2018/04/11/294876

18. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112–21.

19. Davies NM, Dickson M, Smith GD, Van Den Berg GJ, Windmeijer F. The causal effects of education on health outcomes in the UK Biobank. Nat Hum Behav. 2018;2(2):117–25.

20. Davies NM, Holmes M V., Davey Smith G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ. 2018;362.

21. Holmes M V, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. BMJ [Internet]. 2014;349(5):g4164. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25011450%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4091648

22. Stockley CS. The relationships between alcohol, wine and cardiovascular diseases –A review. Nutr Aging. 2016;3(2–4):55–88.

23. Lawlor DA, Harbord RM, Timpson NJ, Lowe GDO, Rumley A, Gaunt TR, et al. The association of C-reactive protein and CRP genotype with coronary heart disease: Findings from five studies with 4,610 cases amongst 18,637 participants. PLoS One. 2008;3(8).

24. Dixon P, Hollingworth W, Harrison S, Davies NM, Smith GD. The causal effect of adiposity on hospital costs: Mendelian Randomisation analysis of over 300,000 individuals from the UK Biobank. bioRxiv [Internet]. 2019;589820. Available from: https://www.biorxiv.org/content/10.1101/589820v1.full

25. Lawlor DA, Tilling K, Smith GD. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45(6):1866–86.

26. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. Int J Epidemiol. 2010;39(2):417–20.

27. Howards PP. An overview of confounding. Part 1: the concept and how to address it. Vol. 97, Acta Obstetricia et Gynecologica Scandinavica. 2018. p. 394–9.

28. Shastry BS. SNPs: impact on gene function and phenotype. Vol. 578, Methods in molecular biology (Clifton, N.J.). 2009. p. 3–22.

29. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol. 2012;8(12).

30. Burgess S, Thompson SG. Interpreting findings from Mendelian randomisation using the MR-Egger method. Eur J Epidemiol. 2017;32(5):377–89.

31. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Vol. 19, Nature Reviews Genetics. 2018. p. 581–90.

32. Lyall DM, Celis-Morales C, Ward J, Iliodromiti S, Anderson JJ, Gill JMR, et al. Association of body mass index with cardiometabolic disease in the UK biobank: A mendelian randomisation study. JAMA Cardiol. 2017;2(8):882–9.

33. Howe LD, Kanayalal R, Harrison S, Beaumont RN, Davies AR, Frayling TM, Davies NM, Hughes A, Jones SE, Sassi F, Wood AR, Tyrrell J. Effects of body mass index on relationship status, social contact and socio-economic position: Mendelian randomization and within-sibling study in UK Biobank. Int J Epidemiol. 2019; doi: 10.1093/ije/dyz240.