Mendelian Randomization as Applied to Coronary Heart Disease, Including Recent Advances Incorporating New Technology

Sarah J. Lewis, PhD

What is Mendelian Randomization?
The field of epidemiology has struggled to make headway in determining whether exposures are causal factors for complex diseases, largely because of the problems of confounding, reverse causation, and bias. To overcome the problems inherent in observational studies, epidemiologists have proposed using genetic variants as proxies for exposures.1,2 The idea is to exploit genetic variants that influence exposure propensity or are involved in metabolism, transport, or cellular uptake of the exposure or are otherwise associated with exposure levels. The advantage of using genetic variants is that because of the random assignment of alleles with respect to subsequent lifestyle factors, they are considered independent of factors that may confound epidemiological studies (see limitations of Mendelian randomization in the next section). Thus, Mendelian randomization combines the 2 fields of epidemiology and genetics, but rather than aiming to understand genetic mechanisms leading to disease, the principle is to use a genetic variant as a proxy for an environmental exposure or intermediate phenotype that may be on the causal pathway of the disease in question. This method is not without problems; however, advances in genetics are helping to overcome these limitations and are likely to increase its utility in uncovering risk factors for disease.

Limitations of Epidemiology
The success stories in the field of epidemiology, such as the finding that smoking causes lung cancer3 and more recently the reduction of sudden infant death syndrome by the identification of risk factors for this condition,4 are often overshadowed by the plethora of false-positive findings reported in the press and then later refuted. The problem is that although associations may have been observed, association is not causation; only randomized controlled trials (RCTs) are able to test causation, and many exposures found to be associated with disease in observational studies are found not to be causal factors when tested in RCTs. For example, case-control and cohort studies have consistently found that vitamin E is associated with a reduced risk of coronary heart disease (CHD).5,6 However, a meta-analysis of 7 RCTs showed that vitamin E had no effect on any cardiovascular event (odds ratio, 0.98; 95% CI, 0.94 to 1.03).7 The main reasons for the disagreement in findings between observational epidemiology and RCTs are confounding, reverse causation, bias, and measurement error.

Confounding
A confounding factor in an epidemiological study is a factor that is associated with both the risk factor of interest and the outcome and results in an erroneous association between the exposure and outcome. Conventional epidemiology is prone to confounding because lifestyle factors are highly correlated with each other.8 Most exposures do not occur in isolation; for example, individuals who have a low vitamin E intake tend to have a higher body mass index (BMI), tend to drink more alcohol, and are more likely to smoke; they also tend to be from a manual social class; and they may have a number of adverse socioeconomic and behavioral risk factors that would make them more susceptible to CHD and other complex diseases.9

Methods for accounting for potential confounders by stratification and statistical adjustment exist, but when adjusting for confounders, one is assuming that those confounders have been adequately measured and all potential confounders have been included in the model; this assumption is likely to be unrealistic. To illustrate the degree to which lifestyle factors cluster together, Davey Smith et al8 analyzed data on behavioral, socioeconomic, and physiological factors from a cohort study of middle-aged women. Among 96 potential risk factors, 45% of all pairwise associations between the factors were significant at the P<0.01 level.

Reverse Causation
A further problem in observational studies is the inability to assign direction to observed associations or temporal sequence to events. For instance, moderate alcohol intake is thought to protect against CHD10; however, it is not clear which appears first: is it that a small amount of alcohol protects against CHD risk, certainly the effect of alcohol on increasing high-density lipoprotein cholesterol levels would...
suggest that this is the case, or is it that individuals with underlying illness such as high blood pressure or atherosclerosis self-medicate by abstaining from alcohol, which gives the apparent association of a reduced risk among moderate drinkers compared with abstainers?

Increased C-reactive protein (CRP) levels have been found among patients with CHD relative to controls, even when prospectively measured in cohort studies, which has provoked much interest in CRP as a potential therapeutic target, but this association is not as it seems. Reassessment of the causal role of CRP in a Mendelian randomization framework seems to suggest that increased CRP does not cause heart disease, but that CRP may be increased as a result of inflammatory processes that occur in preclinical CHD. In the field of epidemiology, the phenomenon whereby the disease causes the hypothesized risk factor and not vice versa is known as reverse causation.

Bias

Reporting bias, interviewer bias, and response bias, along with other forms of bias, are further explanations for associations found in observational studies that are then not substantiated in RCTs. Individuals with a disease will often report their exposures differently to the general population because they are likely to be particularly sensitive to anything that may have caused them to become ill and, relative to controls, may overreport exposures (reporting bias). One example of this may be proximity to power lines among parents of leukemia cases, with parents of cases being more likely to be aware of surrounding power lines than are parents of controls. Similarly, interviewers may ask questions differently if they are aware that the interviewee is a case rather than a control (interviewer bias). Individuals may even be more likely to take part in a study if they have been exposed to a particular factor that they believe has caused their illness (response bias).

Measurement Error

Finally, observational studies are often unable to measure exposures accurately, and this measurement error may lead to false claims of associations between exposure and disease. In fact, “noise” in the measurement of exposure in cases and controls is more likely to lead to attenuation of exposure-disease associations, which could mean that risk factors are not identified. For example, many studies of dietary intake use a food frequency questionnaire, and it has been shown that these instruments are subject to substantial measurement error. This is due to a combination of misreporting of food intake by study subjects, imprecise questions on food intake, and measurement error in the tool that converts food intake into nutrient levels. Indeed, a recent editorial questioned whether it was “time to abandon the food frequency questionnaire?” This may explain why, after thousands of studies, it is still not clear which nutrients are risk and protective factors for common diseases.

Why Not Just Carry Out an RCT?

Because observational studies are so problematic, it would be preferable to abandon these in favor of the “gold standard,” which is the RCT. RCTs are free from confounding because the treatment or exposure is allocated at random and is therefore not associated with factors that confound observational studies. Similarly, blinding in an RCT can remove bias, both on the part of the participant and of the person assessing the outcome. However, for many questions, RCTs are simply not feasible or would be prohibitively expensive. An RCT of whether taking folic acid supplements during pregnancy affects CHD risk in the offspring would need to have a long follow-up, and it would also be unethical to randomize some women to a placebo because folic acid is known to protect against neural tube defects.

Mendelian randomization has been proposed as a method akin to an RCT. Within a given population, the inheritance of a particular allele at a given locus is independent of all other alleles at other loci except for those that are in LD with it, as described by Gregor Mendel in what was later to be referred to as his law of independent assortment, which states that traits are transmitted to offspring independently of one another. However, the important feature of genotypes with respect to epidemiology is that the inheritance of alleles is assigned at conception and generally not associated with lifestyle factors that may confound an observational study. In an RCT, folic acid would be assigned at random and would therefore not be associated with smoking, social class, alcohol intake, etc because these factors would, on average, be equally distributed between the intervention and control groups. The Table shows that genotype at locus C677T in the methylene tetrahydrofolate reductase (MTHFR) gene, which affects the metabolism of folate and resulting homocysteine levels, is not associated with lifestyle factors that may confound an association between folate intake and disease risk.

Aside from random distribution of confounders with respect to the exposure of interest, other parallels can be drawn between RCTs and Mendelian randomization as described in the following sections.

Concealment

Allocation of exposure or treatment in an RCT is preferably performed without previous knowledge of the allocation sequence; ideally, the randomization process would be unpredictable to prevent the occurrence of selection bias. In the case of Mendelian randomization, the process of allocation of alleles is a random process of nature and is therefore concealed.

Blinding

Blinding refers to the knowledge of treatment received among patients, healthcare professionals, and outcome assessors during the trial. As many individuals involved as possible should be blind to the treatment allocation to prevent reporting and assessment bias. In Mendelian randomization, individuals and healthcare professionals are not normally aware of an individual’s genotype, with the exception of genotypes that have dramatic phenotypic effects, such as the aldehyde dehydrogenase-2 (ALDH2) genotype and alcohol intake (see later).
Intention to Treat
In an RCT, individuals should be analyzed according to the group to which they were assigned, regardless of whether they complied with the treatment or changed treatment groups. In Mendelian randomization, individuals are analyzed according to their assigned genotype regardless of their actual exposure. One common criticism of the method is that not all individuals with a particular genotype have the characteristic phenotype; for example, not all individuals with the MTHFR C677T CC genotype have low homocysteine levels, which is a trait associated with this phenotype. This will depend on an individual’s folate intake and other factors. However, a group of individuals with the MTHFR C677T CC genotype will have lower average homocysteine levels than a group of TT individuals (Figure 1).

Similarly, within a group of individuals assigned to take folic acid supplements, some individuals will be noncompliers, in which case they will have low folate and high homocysteine levels relative to the rest of their group, but on average, homocysteine levels among a group allocated to take folic acid supplements will be lower than among a group allocated to a placebo.

Background to This Approach
In 1986, suspecting that observational reports on the apparent dangers of low cholesterol levels for cancer risk were subject to confounding or reverse causation (eg, occult tumors causing lowering of cholesterol levels in seriously ill patients), Katan postulated that natural variation in the apolipoprotein E (APOE) gene, which is strongly associated with cholesterol levels, could be used in the assessment of relations between cholesterol and outcome.

Two separate polymorphic loci in the APOE gene produce 3 independent alleles coding for the major isoforms APO E-2, E-3, and E-4. The different isoforms of apo E are associated with total cholesterol levels as follows: E4 > E3 > E2. A comparison of apo E phenotypes in patients with cancer with those in matched controls might thus shed light on the relation between low cholesterol and cancer. If it is causal, then the E-2 isoform should be more common among patients and E-3 and E-4, more common among controls. Alternatively, an equal distribution of apo E phenotypes among cases and controls would suggest that the association between low cholesterol and cancer is not causal.

Before Katan’s clear description of the principle of using genotypes as proxies for environmental modifiable risk factors, others had used the basic principle. The term “Mendelian randomization” had been proposed with respect to a different genetically informed design for obtaining causal evidence without experimenter-controlled randomization and was later used to describe this principle in the observational studies setting. The concept was explicated at length in 2 review articles on Mendelian randomization in 2003 and 2004, which presented several examples along with categories of inference and some limitations of this approach. This method has since been adopted by many epidemiologists and has led to the following findings: exposure to high levels of glucose in utero increases birth weight and CRP does not cause symptoms of the metabolic syndrome, and differences in fibrinogen concentrations are not a major determinant of coronary disease risk.

Biological Underpinning—Genes as Instruments
Mendelian randomization relies on the random segregation of homologous chromosomes at meiosis, the apparently random action of chromosomal recombination by sexual reproduci-
tion, and the independence of this process with respect to classical confounding factors. In addition, because genotype is determined at conception, it is not susceptible to reverse causation. Genetic variation associated with risk factors of interest may be considered as “better measures” or “proxy measures” of environmental risk. In light of this, genetic polymorphisms may be termed “instruments.”32,33 The use of variables that act as proxy measures for risk factors of interest, which are not subject to confounding, was first described in the field of econometrics. This “instrumental variable” approach was seen to be of value in assessing the causal relation between risk factors and outcomes of interest.

**Categories of Inference**

Critical to the assumptions made in Mendelian randomization and the inferences that can be drawn from the result of experiments undertaken in this framework are the properties of these so-called “instrumental variables” and the causal pathways along which they are seen to operate. Some categories of inference are described next, but these are not meant to be exhaustive and have been described in detail elsewhere.1,2 Therefore, an abridged version is presented here.

**Exposure Propensity**

Genetic variants may influence the tendency to become exposed; in this case, it is possible to define high- and low-exposure groups based on genotype. For example, a common genetic variant in ALDH2, which is present in some Asian populations, is strongly associated with alcohol intake and can be used to define alcohol exposure groups.34

**ALDH2 Genotype and Blood Pressure**

Many epidemiological studies have reported a J-shaped curve between alcohol intake and blood pressure, with moderate drinkers having a lower blood pressure than teetotallers.1,35 However, perhaps individuals with known high blood pressure or who are at risk of high blood pressure have stopped drinking alcohol (reverse causation) or alternatively, perhaps those individuals who drink a single glass of wine a day with their dinner also display moderation in other lifestyle factors, which leads to a low risk of high blood pressure (confounding).

Japanese and some other Asian populations have a flushing response after alcohol ingestion; for some, the presentation of symptoms (dizziness, headaches, and nausea) in response to alcohol is so severe that they simply do not drink. This phenotype is a result of a polymorphism in ALDH2.36 On drinking alcohol, it is metabolized initially to acetaldehyde by alcohol dehydrogenase enzymes and is then further metabolized to acetic acid. The ALDH2 gene codes for the principle enzyme involved in the metabolism of acetaldehyde to acetic acid. Individuals with the ALDH2 *2*2 genotype do not produce active acetaldehyde dehydrogenase and are unable to clear acetaldehyde from the body.36,37 This leads to the symptoms listed above, and individuals with the *2*2 genotype tend not to drink, with heterozygotes drinking an intermediate amount. This genetic variant therefore defines 3 exposure groups with respect to alcohol intake levels (Figure 2).

The ALDH2 genotype provides a natural experiment in which to investigate the association between alcohol and blood pressure because the genotype is not associated with smoking, sex, age, and other factors that may influence blood pressure.38

We investigated the association between this genotype and blood pressure in a meta-analysis of published studies.34 We found that men with the heterozygote genotype who tended to be moderate drinkers had significantly higher blood pressure than did those with the *2*2 genotypes (who were, on the whole, teetotallers). In addition, *1*1 homozygotes (who drank on average 2 to 3 units per day) had blood pressure levels that were 7.5 mm Hg greater than those of the *2*2 individuals. Women who tended not to drink regardless of genotype did not show an association between genotype and blood pressure, which confirms that the effect of ALDH2 on blood pressure is mediated through alcohol intake. This effect of alcohol on blood pressure was much greater than had previously been shown in epidemiological studies and was equivalent to the effect of antihypertensive drugs on blood pressure (Figure 3).

**Intermediate Phenotypes**

An intermediate phenotype is a measurable biological effect that is not the outcome of interest but may lead to the outcome of interest. For example, there are many biomarkers that are associated with CHD risk, such as lipid and lipoprotein particles, proteins involved in inflammation and coagulation, and metabolites and markers of oxidative stress.42 Genetic variants that are very strongly associated with the aforementioned intermediate phenotypes, such as the single-nucleotide polymorphism (SNP) rs6046 in the F7 gene, which explains \( \approx 10\% \) of the variance in circulating blood factor VII levels (probability value for association with FVII = \( \approx 9.2 \times 10^{-140} \)), can be used to delineate causal pathways and identify targets for treatment.42 A minor allele of an apolipoprotein (a) variant (rs3798220) has been reported to be strongly associated with increased plasma lipoprotein(a) [Lp(a)] and with increased cardiovascular disease risk, suggesting that this may be a suitable target for prevention of cardiovascular disease.43,44

**Intergenerational Effects**

Mother’s genotype influences the intrauterine environment to a certain extent, and genotypes that mimic environmental exposures can be used to determine whether early exposure to modifiable environmental factors influences offspring pheno-
If intrauterine environment is a risk factor for disease, one would expect to find an association with the mother’s but not with the father’s genotype. For example, maternal glucose levels are thought to affect her offspring’s birth weight, although associations between maternal glucose levels and offspring birth weight could be heavily confounded. The presence of the A allele of a common variant in the glucokinase gene has been found to be associated with an increase in fasting plasma glucose among 755 pregnant women (0.075 mmol/L, \( P = 0.003 \)) relative to the GG genotype at this locus. Conditioning on offspring genotype, the A allele in the mother was associated with a 64-g (95% CI, 25 to 102 g) increase in offspring birth weight (\( P = 0.001 \)). Freathy et al have since replicated this finding in a larger study, suggesting that fasting glucose levels in the mother do influence the offspring’s birth weight. This association is unlikely to have arisen by confounding as a result of the mother’s diet and other lifestyle factors.

### Limitations of Mendelian Randomization

Although this approach offers an exciting alternative to traditional observational epidemiology, it is not without limitations. Many of the limitations in Mendelian randomization studies are common to all genetic association studies; however, technological advances, which have increased knowledge in the field of genetics, are improving the utility of Mendelian randomization and minimizing the impact of some of the limitations. The limitations, the extent to which they prohibit the use of this method, and the potential to overcome these limitations with advances in genetics are discussed next.

### Lack of Suitable Polymorphisms and Failure to Establish Reliable Gene-Exposure Polymorphisms

Until recently, genetic association studies were not particularly fruitful in identifying genes for disease, with many associations failing to replicate. This was largely due to a combination of extremely underpowered studies, multiple testing, and publication bias (positive findings are more likely to be published than null findings). The problem of genetic association studies failing to replicate has a knock-on effect for Mendelian randomization because this relies on established associations. However, although genotype-disease associations are often found to be false-positives, genotype–intermediate phenotype associations have tended to be more consistent, possibly because these are often reported as an aside to the main outcome and therefore may be less prone to
publication bias, and knowledge of biological pathways means that genes that code directly for or are involved in the expression or metabolism of the exposure or intermediate phenotype can be identified. In addition, genetic variants tend to have larger effects on intermediate phenotypes than on complex disease risk, which means that even small studies have sufficient power to detect associations. Thus, the small number of instruments that are established, eg, ALDH2 and MTHFR C677T, tend to be well characterized.

Genome-wide association studies are proving extremely fruitful in terms of finding disease genes. Many investigators with genome-wide data are now turning their attention to uncovering genes that are associated with exposures and intermediate phenotypes. Genetic variants that are strongly correlated with lipids, omega-3 fatty acids, and vitamin B12 have been uncovered through this method. The FTO gene, which was first identified in genome-wide association studies of diabetes, is strongly correlated with BMI and as such can be used to determine whether BMI is a causal factor in disease. This gene has been used as an instrumental variable to show that BMI is causally related to blood pressure, metabolic traits, adult carotid intima-media thickness, and other atherosclerosis risk factors.

Confounding in Mendelian Randomization

The main impetus for using Mendelian randomization in the epidemiological field is to avoid problems of confounding. However, in certain circumstances, confounding can also occur with this method.

Confounding by Population Structure

Allele frequencies at many common polymorphisms differ widely by population. At the extreme end of the scale, certain alleles may only be present in selected populations, such as the ALDH2 *2 allele, which is present at a frequency of ≈0.35 in Japanese and other Southeast Asian populations but is virtually nonexistent in individuals of European ethnicity. Because disease frequencies also differ by population, spurious associations between genotype and disease can arise if population structure is not adequately controlled for in the analysis. However, providing population structure is controlled adequately, this should not prohibit the use of this technique. In addition, where genome-wide genotype data are available, as is increasingly the case, such data can be used to estimate principal component vectors for hidden population stratification, and these are then included as covariates in a linear regression model. This method effectively corrects for residual population stratification.

Confounding by LD

A further way in which confounding could arise with this method is if another polymorphism, which is in close proximity (and in LD) to the variant of interest, is causing disease through another pathway. For instance, in the ALDH2 and blood pressure analysis, it was assumed that the association between the ALDH2 genotype and blood pressure was due to the effect of this genotype on alcohol intake. However, if this polymorphism is in LD with another polymorphism that influences blood pressure, by affecting clotting factors, for instance, then the assumption that alcohol intake influences blood pressure would be violated. In this case, an investigation of this polymorphism in a population (women) in which there is no association with the exposure of interest provided confirmation that the association with genotype was through alcohol intake. An understanding of the underlying biology will generally help to overcome this problem.

Confounding by Pleiotropy

Often, genes act on multiple pathways and may therefore be associated with multiple exposures or intermediate phenotypes, especially those that act as transcription factors for other genes. In addition, a single gene can code for more than 1 protein because of alternative protein splicing. For instance, the G allele at rs6564851, near the β-carotene 15,15'-monooxygenase 1 (BCMO1) gene, has been found to be associated with higher β-carotene ($P=1.6\times10^{-24}$). Ideally, this could be used to determine whether β-carotene protects against CHD and other diseases. However, this same allele is also associated with lower lycopene (0.003), zeaxanthin ($P=1.3\times10^{-5}$), and lutein ($P=7.3\times10^{-15}$) levels, with effect sizes ranging from 0.10 to 0.28 SDs per allele. Perhaps the effects on lycopene, zeaxanthin, and lutein are due to the antagonistic behavior of β-carotene on other antioxidants and therefore downstream of β-carotene. In which case, assumptions made about the role of β-carotene using this genotype would be valid. However, if this polymorphism is influencing these biomarkers by other pathways that are independent of β-carotene, then inferences drawn on the effects of β-carotene may be invalid. Again, this problem can be avoided with knowledge of the underlying biology.

The cleanest instruments for this type of analysis are cis-acting single-nucleotide polymorphisms, such as in the CRP example. These cis-acting variants in genes encoding a quantitative trait provide highly specific instruments with which to investigate the causal effects of the encoded protein. For example, individuals homozygous for the rare variant allele rs6046 in the F7 gene exhibited a 2.15-SD higher blood factor VII level than those who were homozygous for the common allele, which is approximately equal to the difference between the top and bottom tertiles of the factor VII distribution. Any associations between this variant and other biomarkers of CHD risk must represent pathways on which this protein is acting.

Another solution to the problems of confounding by population structure or LD and pleiotropy in Mendelian randomization is to use multiple instruments, or genes, that influence the exposure or intermediate phenotype through different pathways. An association with 1 polymorphism could have arisen by chance or confounding, but associations with more than 1 polymorphism in different genes marking the same exposure are unlikely unless the exposure is causing the disease. This is becoming increasingly possible as more genetic variants are uncovered for exposures or intermediate phenotypes of interest.

Developmental Stability

The final limitation of Mendelian randomization is that of developmental compensation. For the most part, it is not
clear to what extent adaptation to genetic mutations occurs and whether polymorphisms that result in loss of function of a particular protein are compensated for. A genetic variant could be associated with a difference in the intermediate phenotype, but if this perturbation of the intermediate phenotype has been compensated for in the canalization/developmental compensation framework, this difference may not be related to disease risk. Evidence that this occurs comes from knockout animals, where removal of the gene has a lesser effect than predicted, given what is known of the function of the gene. It is not yet clear to what extent this will influence Mendelian randomization, particularly if the variant only has a minor impact on exposure levels. Although exposures, such as alcohol, that appear in adulthood are unlikely to have been compensated for during development, in many instances, this can be verified by performing gene-disease analysis where the related exposure is known to cause disease. For example, alcohol is known to cause esophageal cancer. Similarly, the ALDH2 genotype is associated with esophageal cancer in the expected direction and to the extent predicted by the effect of this genotype on alcohol exposure. This suggests that the ALDH2 genotype is a good proxy for exposure to alcohol and that canalization has not occurred in this pathway. Another way of determining whether canalization has occurred could be to look for epigenetic markers and determine whether there are differences in methylation patterns of other genes by genotype, because any compensatory mechanism is likely to be epigenetic. However, epigenetic markers are context and time dependent, and further work in this field is required before this approach can be used.

**Frequently Asked Questions**

My variant only explains a small amount of variance in exposure; can I use this in Mendelian randomization?

The size of the effect of the genetic variant does not matter, providing your study is adequately powered to detect an effect between the intermediate phenotype or exposure and outcome, given the size of effect of the genotype on intermediate phenotype and the expected size of the effect of the intermediate phenotype on disease. However, one common mistake is to conclude that because the instrument is not associated with disease, the corresponding exposure is also not associated, whereas the real problem is often a lack of power to detect an effect. It is only possible to rule out an effect size that falls outside your CIs.56

I don’t have intermediate phenotype/exposure measures, so does this mean that I can’t do a Mendelian randomization analysis?

To draw inferences using a Mendelian randomization framework it is not always necessary to have measured the intermediate phenotype or exposure in your own study. You can draw on knowledge from other studies to make assumptions relating to your own study. Bear in mind that if you are using data from other studies, you should check that they have been performed in a population comparable to that of your own study. (Genetic variants may be associated with exposures because they are in LD with the actual functional variant. Because LD structures differ by population, the same variant may not be associated with exposure in your population.)

Similarly, cultural reasons may exist that explain different associations between exposure and genotype in different populations. For instance, the lactase persistence genotype, which determines whether individuals can digest lactose in cow’s milk, is not associated with milk intake in all populations. This may be due to the cultural acceptance of not drinking milk (it may be more acceptable in populations where a large proportion of individuals are lactose intolerant but not in populations where the majority are lactose persistent).

What are the public health implications of my findings, given that I am not likely to alter genotype to reduce disease risk?

The FTO gene variant is only associated with a modest change in atherosclerotic risk factors and is therefore unlikely to ever be used directly to reduce atherosclerosis by genetic engineering. This is because this genotype only explains a small proportion of variance in BMI; people who are genetically susceptible do not always become overweight, and those who are not susceptible sometimes do. The finding that this gene is associated with atherosclerotic risk factors suggests that being overweight and obese is a causal factor for atherosclerosis. Public health programs aimed at reducing atherosclerosis can therefore target obesity prevention, which will benefit everyone, not just those with the FTO risk allele. The aim of Mendelian randomization is to identify “modifiable exposures” for disease for the benefit of the whole population, not to identify those with slightly increased susceptibility because of their genetic makeup.

Are all variants within a gene (that affect a trait) equally capable of being instruments for Mendelian randomization, or can this only be done with variants that affect protein levels as opposed to function?

Any variant can be used for Mendelian randomization, provided that it is consistently associated with the intermediate phenotype or exposure in question and that the results are interpreted in this context. The rs6046 variant in FTO affects protein levels, whereas the C677T variant in MTHFR produces a protein with lower levels of activity at 37°C, which affects the efficiency of the folate metabolic pathway and ultimately, levels of downstream metabolites.

**Acknowledgments**

I thank George Davey Smith, Nic Timpson, Beate Glaser, and Luisa Zuccolo for reading and commenting on earlier versions of this review.

**Disclosures**

None.

**References**


Key Words: Mendelian randomization ■ epidemiology ■ genetics ■ polymorphisms ■ environment ■ confounding