Using Mendelian randomisation to assess the potential benefit of a clinical intervention

Stephen Burgess  Adam Butterworth  Anders Malarstig
Simon G. Thompson
August 28, 2012

Stephen Burgess and colleagues investigate the relevance of associations between genetic variants and a modifiable risk factor to judge the potential therapeutic benefit of a clinical intervention.

Standfirst: Mendelian randomisation is a technique for assessing causal associations in observational data. Genetic variants associated with the risk factor of interest are regarded in a similar way to random assignment in a clinical trial. A difference in the disease outcome between genetically-defined groups is interpreted as being causally due to the risk factor, under the assumption that all possible confounding variables are balanced between the groups. However, the difference in the risk factor due to the genetic variation is materially distinct from the change due to any proposed therapeutic intervention, and so may affect the outcome differently. Consequently, it may be misleading to generalise the magnitude of a Mendelian randomisation estimate to the effect of a potential intervention on the risk factor in practice. Awareness of the limitations of such estimates is important for the use of Mendelian randomisation in target-based drug development. (146 words)

What is Mendelian randomisation?

If epidemiologists are compared to fishermen, causality is the big fish. It is elusive to find, difficult to catch, and claims to have measured it are often exaggerated. But, despite the challenge, demonstration of causal relationships remains a central aim of epidemiological enquiry. Mendelian randomisation is becoming a commonly used technique to make assessment of causality possible from observational data. For example, it has recently strengthened the case for a causal role in coronary disease of lipoprotein(a), and weakened the case of C-reactive protein.

To perform Mendelian randomisation, a genetic variant is sought which has three key features. First, it is associated with the risk factor of interest. Secondly, it divides the observed population into groups similar to arms in a randomised trial, which do
not systematically differ with respect to any confounding variable. This ensures that any difference in the outcome is due to the genetic variant. Thirdly, it affects the outcome only via the risk factor of interest and not by other biological pathways. Provided these key features hold, a causal association of the risk factor on the outcome is inferred. The genetic variant acts as a proxy for the risk factor, and the random allocation of genes at conception is exploited as a natural experiment to demonstrate causation. Figure 1 shows the correspondence between Mendelian randomisation and a randomised trial.

Figure 1: Comparison of Mendelian randomisation and a randomised trial

For example, the genetic variant rs11206510 is associated with both low-density lipoprotein cholesterol and coronary heart disease, as shown in a recent study. Each additional copy of the C allele is associated with a 2.5% reduction in low-density lipoprotein cholesterol and an odds ratio for coronary heart disease of 0.93. Under the Mendelian randomisation assumptions that these associations are not confounded and that the genetic association with the disease is entirely mediated by the risk factor, we estimate that a 2.5% reduction in low-density lipoprotein cholesterol leads to a 7% reduction in coronary heart disease.

What factors affect the validity of Mendelian randomisation studies?

From the first discussions of Mendelian randomisation, researchers have emphasised that the assumptions leading to the assertion of causal association may be invalid for many genetic variants. Violations in the assumptions of no direct association between the genetic variant and either the outcome or any confounding risk factor may occur for several reasons. These reasons include the association of the variant with multiple risk factors (pleiotropy), the association between the variant and other genetic variants
(linkage disequilibrium), and the presence of genetic differences between possibly hidden subgroups in the population under investigation (population stratification). For example, in a North American cohort, a variant may be associated with type 2 diabetes due to increased prevalence of the variant amongst those of Native American descent, who are known to have a greater incidence of the disease. The genetic association may be driven not specifically by a single risk factor, but by a range of factors associated the difference in ethnic background. Such violations of internal validity can lead to misleading conclusions.

An aspect of validity which has received less attention is the issue of external validity. If the assumptions about the genetic variant are true and a valid estimate is made which corresponds to a causal association, can this estimate be generalised to the effect of a clinical intervention? For example, is the estimate of lowered risk derived from considering genetically reduced levels of cholesterol the same as the lowered risk conferred by an intervention that reduces levels of cholesterol?

**Reasons why Mendelian randomisation may give a different estimate to an intervention**

Mendelian randomisation is different from a randomised trial in a fundamental way. In a randomised trial, the intervention applied to the treatment group is usually the intervention which is proposed in clinical practice. In Mendelian randomisation, the “intervention” leading to differences between the groups within the study is the presence of a genetic variant. The question of external validity is whether the causal effect due to the change in the risk factor as a result of the presence of the genetic variant is similar to the causal effect due to the proposed intervention on the risk factor. Aside from those due to differences in the study population, there are several reasons why these effects may be unequal.

(i) **Time-scale and developmental compensation**

The presence or absence of the genetic variant in an individual is determined at conception. This means that the Mendelian randomisation estimate represents the result of a life-long difference in the risk factor between the groups. An intervention in levels of a risk factor for coronary heart disease (for example) may have limited benefit as some stages of atherosclerosis may be irreversible. There may be no intervention on the risk factor in a mature cohort which can imitate the genetic effect. The same would be true if the disease progression depends on a developmental phase at an particular stage of life.

For some risk factors, an individual may develop compensatory mechanisms (canalisation) in response to elevated (or lowered) levels of the risk factor. This has been seen in knockout studies, where deletion of a particular gene often does not have the profound effect expected. This is because alternative pathways are developed as a compensatory mechanism to circumvent the missing gene. For example, previous studies of interleukin-1 knockout mice have suggested that other inflammatory responses...
(for example, tumour necrosis factor-alpha levels) might be elevated to compensate for the loss of inflammatory signalling from the interleukin-1 pathway\textsuperscript{11}.

(ii) Usual versus pathological levels

Disease risk often depends primarily on the average or “usual” levels of a risk factor. Mendelian randomisation has a particular role to play here, as genetic variants would be expected to affect these average levels. It is however plausible that long-term elevated average levels of a risk factor do not affect disease risk, but acute response of the risk factor does\textsuperscript{8}. The efficacy of short-term targeted interventions on pathological levels of a risk factor may not be validly assessed by Mendelian randomisation.

For example, genetic variants which are associated with usual levels of C-reactive protein have been used to assess the causal association of long-term elevated average levels of C-reactive protein on cardiovascular risk\textsuperscript{3}. Although the causal association between C-reactive protein and cardiovascular risk appears to be null, this does not preclude the efficacy of a therapeutic intervention on acute levels of C-reactive protein, which is better assessed by \textit{in vivo} studies\textsuperscript{12}.

(iii) Extrapolation of small differences

The change in a risk factor due to genetic variants is generally small. For an intervention lowering (or raising) the risk factor uniformly by a small amount for everyone in the population, a Mendelian randomisation study may provide a relevant estimate of the effect of the intervention. However, if the proposed intervention in the risk factor is more substantial, then the Mendelian randomisation estimate of its effect relies on extrapolation.

For example, the effect of statin use (inhibition of 3-hydroxy-3-methylglutaryl-coenzyme A reductase) on low-density lipoprotein cholesterol levels is several times larger than the association of low-density lipoprotein cholesterol levels with variants in the \textit{HMGCR} gene. Extrapolation of the genetic effect relying on a linear assumption for the effect of the risk factor on the outcome may not be valid.

(iv) Different pathways of genetic and intervention effects

The genetic variant and the proposed intervention will not, in general, have the same specific effect on the risk factor. This situation is similar to that of differences between drugs which act on different mechanisms but influence the same mediating risk factor. The genetic change in the risk factor may be associated with another variable, as in the case of a variant in the \textit{FTO} gene associated with obesity\textsuperscript{13}. The effect of \textit{FTO} on obesity is not direct; rather the genetic variant affects satiety, which in turn affects obesity\textsuperscript{14}. An intervention on obesity which is not based on reducing food intake may have a different effect on the outcome to a Mendelian randomisation study. Even when both effects are specifically targeted on the risk factor, it may be that they are on different biological, biochemical or physiological pathways, and so the genetic and clinical changes in risk factor may affect the outcome to different extents.
Cholesterol and coronary heart disease

We give an example to illustrate the differences between Mendelian randomisation estimates and those from other approaches. Coronary heart disease is the result of a build-up of atheromatous plaques in the coronary arteries. A major component of such plaques is cholesterol, and low-density lipoprotein cholesterol is an established causal risk factor for coronary heart disease. We assess the association between low-density lipoprotein cholesterol and coronary heart disease from Mendelian randomisation, and from randomised trials where the clinical intervention for low-density lipoprotein cholesterol lowering is statin use.

We consider 5 genetic variants from a meta-analysis of genome-wide association studies which are associated with low-density lipoprotein cholesterol, but not with high-density lipoprotein cholesterol nor triglycerides. Table 1 gives the estimates of association of each genetic variant with log-transformed low-density lipoprotein cholesterol and risk of coronary heart disease, and Mendelian randomisation estimates using each genetic variant of the causal odds ratio of coronary heart disease per 30% decrease in low-density lipoprotein cholesterol. These odds ratios range from 0.27 to 0.45. We note that this relies on a between 8 and 20-fold extrapolation of the genetic effects on the risk factor.

<table>
<thead>
<tr>
<th>Genetic variant (relevant gene)</th>
<th>Per allele change in log-transformed low-density lipoprotein cholesterol (standard error)</th>
<th>Per allele odds ratio of coronary heart disease (95% confidence interval)</th>
<th>Odds ratio of coronary heart disease per 30% decrease in low-density lipoprotein cholesterol</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11206510 (PCSK9)</td>
<td>−0.026 (0.004)</td>
<td>0.93 (0.88 to 0.99)</td>
<td>0.40 (0.15 to 0.85)</td>
<td>0.357</td>
</tr>
<tr>
<td>rs660240 (SORT1)</td>
<td>−0.044 (0.004)</td>
<td>0.85 (0.80 to 0.90)</td>
<td>0.27 (0.15 to 0.44)</td>
<td>0.357</td>
</tr>
<tr>
<td>rs515135 (APOB)</td>
<td>−0.038 (0.004)</td>
<td>0.90 (0.85 to 0.96)</td>
<td>0.37 (0.19 to 0.66)</td>
<td>0.357</td>
</tr>
<tr>
<td>rs12916 (HMGCR)</td>
<td>−0.023 (0.003)</td>
<td>0.94 (0.90 to 0.99)</td>
<td>0.38 (0.16 to 0.80)</td>
<td>0.357</td>
</tr>
<tr>
<td>rs2738459 (LDLR)</td>
<td>−0.018 (0.004)</td>
<td>0.96 (0.89 to 1.03)</td>
<td>0.45 (0.07 to 1.95)</td>
<td>0.357</td>
</tr>
</tbody>
</table>

Table 1: Association of five genetic variants with log-transformed low-density lipoprotein cholesterol and coronary heart disease risk taken from Waterworth et al. Causal estimates (with 95% confidence intervals) of odds ratio for 30% reduction in low-density lipoprotein cholesterol on coronary heart disease from Mendelian randomisation using each genetic variant in turn.

In comparison, randomised trials of statins have given lower estimates of the benefits of reduced low-density lipoprotein cholesterol levels. A meta-analysis examining the effect of statin use on coronary heart disease, comprising 69 139 participants with 6406 events, gave a relative risk of 0.73 (95% confidence interval 0.70 to 0.77) based on average reduction of around 30% in low-density lipoprotein cholesterol over an average follow-up time of at least 3 years. A more focused meta-analysis examining the effect of statin use for primary disease prevention, comprising 27 969 individuals without previous history of coronary heart disease with 1677 events, gave a similar
relative risk of 0.72 (95% confidence interval 0.65 to 0.79) over 1.5 to 3 years follow up\textsuperscript{17}.

It is known that the effect of statins in reducing coronary heart disease increases over time\textsuperscript{18}. As atherosclerosis is a chronic condition which develops progressively, it is not surprising that the estimates of the effect of the life-long lowering of low-density lipoprotein cholesterol associated with the genetic variants corresponds to a greater proportional change in coronary heart disease risk than the effects of statin usage. The difference between the estimates could also be due to the non-specific effects of statins; however any effects of statins on inflammatory response would further lessen the role of low-density lipoprotein cholesterol, and make the contrast with the genetic effects more extreme.

**Blood pressure and coronary heart disease**

Another example is the association between blood pressure and coronary heart disease. A genetic risk score associated with a 1.6mmHg decrease in systolic blood pressure corresponds to an odds ratio for coronary heart disease of 0.91 (95% confidence interval 0.89 to 0.92)\textsuperscript{19}. Assuming a linear association, this implies an odds ratio of 0.55 (95% confidence interval 0.47 to 0.61) for a 10mmHg decrease in systolic blood pressure, compared to the relative risks from a meta-analysis of 0.78 (95% confidence interval 0.73 to 0.83) in trials and 0.75 (95% confidence interval 0.73 to 0.77) in cohort studies\textsuperscript{20}. Here again, the estimate of the benefit of reducing blood pressure from the genetic variants is much greater than that of the intervention.
Box: Interpreting the result of a Mendelian randomisation study

A Mendelian randomisation study tests whether a risk factor is causally associated with a disease outcome by examining whether there are differences in the outcome between genetically-defined groups with different average levels of the risk factor of interest.

There are three pitfalls in interpreting the result of a Mendelian randomisation study:

**Failure of key assumptions:** The key assumption is that the genetic variant, which is associated with the risk factor, divides the population into groups which are similar to treatment arms in a randomised trial, in that all potential confounding factors are balanced between the groups. This requires lack of pleiotropy of the variant, absence of linkage disequilibrium with other functional variants, and absence of hidden population strata (see text). If any of these conditions do not hold, then estimates from Mendelian randomisation may be misleading.

**Over-interpreting a null finding:** The differences in the risk factor between the genetic groups are usually small compared to its overall variation. A null finding may simply reflect that the small differences between the groups do not result in large enough differences in the outcome to be reliably distinguished from chance differences in a limited sample size. Sample sizes in tens of thousands are required in some cases to provide sufficient power to reliably interpret a null finding\(^8\).

**Over-interpreting a positive finding:** While the Mendelian randomisation hypothesis relates to genetic groups, one aim of a Mendelian randomisation is to inform on the potential effect of a clinical intervention in the risk factor of interest. Qualitative and quantitative differences between the comparison of genetic groups and the proposed intervention mean that the causal effect estimated by Mendelian randomisation may not directly translate into the observed effect on the outcome of modifying the risk factor in practice. (294 words)

Using the Mendelian randomisation paradigm to guide drug discovery

Questions of generalisability of results are especially relevant for the use of Mendelian randomisation in guiding clinical interventions and drug discovery. Genetic evidence can inform the causal role of a risk factor which is being considered as a target for intervention. Association between a relevant genetic variant affecting the risk factor and the outcome may be taken as evidence for the potential efficacy of a drug affecting the risk factor pathway. However, for the reasons given above, absence of
evidence for such an association does not necessarily imply lack of efficacy. Although we may expect Mendelian randomisation in many circumstances to provide a good qualitative indication, the magnitude of the Mendelian randomisation estimate will not necessarily be a reliable guide to the potential benefit of a drug.

In the examples considered, evidence from Mendelian randomisation suggests that low-density lipoprotein cholesterol and blood pressure are appropriate targets for interventions aimed at reducing coronary heart disease. The genetic variants may also suggest particular biochemical pathways for such intervention. In this way, Mendelian randomisation can be used to prioritise risk factors for future pharmacological investigation.

**Prospects for Mendelian randomisation**

Mendelian randomisation is a useful tool for exploring causal relationships between modifiable risk factors and outcomes of interest. It is one of the few epidemiological methodologies that can aid the selection of targets for therapeutic intervention. However, it would be misleading to assume that the estimate from a Mendelian randomisation study gave the definitive answer to the general question of causal relevance of a risk factor. Mendelian randomisation estimates are especially relevant when the effect of interest is that of a long-term population-based intervention. We conclude that, while a Mendelian randomisation approach will generally be qualitatively informative for the direction of effect of a clinical intervention, the genetically derived estimate may not correspond to the magnitude of the effect in practice. (2218 words)

**Key messages:**

- Estimates from Mendelian randomisation represent causal effects of genetically determined differences in a risk factor on a disease outcome
- These estimates are informative for assessing aetiological associations of risk factors and for prioritising targets for pharmaceutical intervention
- However the effects of such interventions may be quantitatively different to those obtained from Mendelian randomisation (53 words)

**Authors**

SB is a post-doctoral researcher working on methods for Mendelian randomisation. AB and AM are genetic epidemiologists, SGT is a statistician. All the authors have a long-standing interest in cardiovascular epidemiology, and specifically in the use of genetic information to guide the development of pharmaceutical research and clinical policy. AM is currently on secondment to the University of Cambridge from Pfizer Pharmaceuticals. SB was the lead author, and the other authors engaged at all stages in the conceptualising and editing of this article. SB is guarantor.
Conflict of interest

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: all authors receive support through grant RG/08/014 from the British Heart Foundation; AM is a current employee of Pfizer Pharmaceuticals; no other relationships or activities that could appear to have influenced the submitted work.

Licence

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd, and its Licensees to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and to exploit all subsidiary rights, as set out in our licence (bmj.com/advice/copyright.shtml).

References


