

## RESEARCH METHODS & REPORTING

### Use of Mendelian randomisation to assess potential benefit of clinical intervention

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. The difference in the risk factor due to the genetic variation, however, is materially distinct from the change due to any proposed therapeutic intervention and so might affect the outcome differently. Consequently, it can 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

Stephen Burgess *statistician*<sup>1</sup>, Adam Butterworth *genetic epidemiologist*<sup>2</sup>, Anders Malarstig *genetic epidemiologist*<sup>3</sup>, Simon G Thompson *statistician*<sup>2</sup>

<sup>1</sup>Strangeways Research Laboratory, Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK ; <sup>2</sup>Department of Public Health and Primary Care, University of Cambridge, Cambridge; <sup>3</sup>Pfizer Global Research and Development, Cambridge

#### What is Mendelian randomisation?

If epidemiologists are compared with 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 relations remains a central aim of epidemiological inquiry. Mendelian randomisation is becoming a commonly used technique to make assessment of causality possible from observational data.<sup>1</sup> For example, in coronary disease it has recently strengthened the case for a causal role of lipoprotein(a)<sup>2</sup> and weakened the case of C reactive protein.<sup>3</sup>

To perform Mendelian randomisation, we look for a genetic variant with three key features. Firstly, 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.<sup>4</sup> This ensures that any difference in the outcome is because of the genetic variant. Thirdly, it affects the outcome only through the risk factor of interest and not by other biological pathways. Provided these key features hold, we can infer a causal association of the risk factor on the outcome.<sup>5</sup> 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 show causation. The figure shows the correspondence between Mendelian randomisation and a randomised trial.<sup>4</sup>

For example, the genetic variant rs11206510 is associated with both low density lipoprotein cholesterol concentrations and coronary heart disease.<sup>6</sup> Each additional copy of the C allele is associated with a 2.5% reduction in low density lipoprotein cholesterol concentration 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 concentration 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 could be invalid for many genetic variants.<sup>7</sup> Violations in the assumptions of no direct association between the genetic variant and either the outcome or any confounding risk factor can 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).<sup>8</sup> For example, in a North American cohort, a variant could be associated with type 2 diabetes because of

increased prevalence of the variant among people of Native American descent, who are known to have a greater incidence of the disease. The genetic association might be driven not specifically by a single risk factor but by a range of factors associated with the difference in ethnic background. Such violations of internal validity can lead to misleading conclusions.<sup>1,5</sup>

An aspect of validity that 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 that 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 concentrations of cholesterol the same as the lowered risk conferred by an intervention that reduces concentrations of cholesterol?

## Reasons why Mendelian randomisation might 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 that 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.<sup>9</sup> 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 resulting from differences in the study population,<sup>10</sup> there are several reasons why these effects might be unequal:

### 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 lifelong difference in the risk factor between the groups. An intervention in levels of a risk factor for coronary heart disease (for example) might have limited benefit because some stages of atherosclerosis might be irreversible. There might be no intervention on the risk factor in a mature cohort that 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.<sup>8</sup>

For some risk factors, an individual might develop compensatory mechanisms (canalisation) in response to increased (or reduced) levels of the risk factor.<sup>9</sup> This has been seen in knockout studies, in which 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.<sup>7</sup> For example, previous studies of interleukin 1 knockout mice have suggested that other inflammatory responses (for example, tumour necrosis factor alpha concentrations) might be increased to compensate for the loss of inflammatory signalling from the interleukin 1 pathway.<sup>11</sup>

### Usual versus pathological levels

Risk of disease 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 plausible, however, that long term increased average levels of a risk factor do not affect risk of disease but acute response of the risk factor does.<sup>8</sup>

The efficacy of short term targeted interventions on pathological levels of a risk factor might not be validly assessed by Mendelian randomisation.

For example, genetic variants that are associated with usual concentrations of C reactive protein have been used to assess the causal association of long term raised average concentrations of C reactive protein on cardiovascular risk.<sup>3</sup> Though there does not seem to be a causal association between C reactive protein and cardiovascular risk, this does not preclude the efficacy of a therapeutic intervention on acute concentrations of C reactive protein, which is better assessed by *in vivo* studies.<sup>12</sup>

### Extrapolation of small differences

The change in a risk factor because of 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 can provide a relevant estimate of the effect of the intervention. If the proposed intervention in the risk factor is more substantial, however, 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 (HMG CoA) reductase) on low density lipoprotein cholesterol concentrations is several times larger than the association of low density lipoprotein cholesterol concentrations with variants in the HMG CoA reductase gene. Extrapolation of the genetic effect relying on a linear assumption for the effect of the risk factor on the outcome might not be valid.

### 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 that act on different mechanisms but influence the same mediating risk factor. The genetic change in the risk factor might be associated with another variable, as in the case of a variant in the FTO gene associated with obesity.<sup>13</sup>

The effect of FTO on obesity is not direct; rather the genetic variant affects satiety, which in turn affects obesity.<sup>14</sup> An intervention on obesity that is not based on reducing food intake might have a different effect on the outcome to a Mendelian randomisation study. Even when both effects are specifically targeted on the risk factor, it could be that they are on different biological, biochemical, or physiological pathways, and so the genetic and clinical changes in risk factor might affect the outcome to different extents.

### Cholesterol and coronary heart disease

We can illustrate the differences between Mendelian randomisation estimates and those from other approaches in the following example. 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 assessed the association between low density lipoprotein cholesterol and coronary heart disease from Mendelian randomisation and from randomised trials in which statins are used as the clinical intervention to lower low density lipoprotein cholesterol concentrations.

We considered five genetic variants from a meta-analysis of genome-wide association studies that are associated with low

density lipoprotein cholesterol but not with high density lipoprotein cholesterol nor triglycerides.<sup>6</sup> The table gives the estimates of association of each genetic variant with log transformed concentrations of low density lipoprotein cholesterol and risk of coronary heart disease, and Mendelian randomisation estimates with each genetic variant of the causal odds ratio of coronary heart disease per 30% decrease in concentration.<sup>5</sup> These odds ratios range from 0.27 to 0.45. This relies on an extrapolation of between eightfold and 20-fold of the genetic effects on the risk factor.

In comparison, randomised trials of statins have reported lower estimates of the benefits of reduced low density lipoprotein cholesterol concentrations. A meta-analysis examining the effect of statin use on coronary heart disease, comprising around 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 concentrations over an average follow-up time of at least three years.<sup>15</sup> A more focused meta-analysis examining the effect of statin use for primary disease prevention, comprising around 27 969 individuals without a 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 three years' follow-up.<sup>16</sup>

The effect of statins in reducing coronary heart disease increases over time.<sup>17</sup> As atherosclerosis is a chronic condition that develops progressively, it is not surprising that the estimates of the effect of the lifelong reduction of low density lipoprotein cholesterol associated with the genetic variants corresponds to a greater proportional change in the risk of coronary heart disease than the effects of statin use. The difference between the estimates could also be caused by the non-specific effects of statins; any effects of statins on inflammatory response, however, 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 1.6 mm Hg 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).<sup>18</sup> Assuming a linear association, this implies an odds ratio of 0.55 (0.47 to 0.61) for a 10 mm Hg decrease compared with the relative risk from a meta-analysis of 0.78 (0.73 to 0.83) in trials and 0.75 (0.73 to 0.77) in cohort studies.<sup>19</sup> Here again, the estimate of the benefit of reducing blood pressure from the genetic variants is much greater than that of the intervention.

## Using the Mendelian randomisation paradigm to guide drug discovery

Questions of generalisability of results are especially relevant when Mendelian randomisation is used in guiding clinical interventions and drug discovery. Genetic evidence can inform the causal role of a risk factor that is being considered as a target for intervention. Association between a relevant genetic variant affecting the risk factor and the outcome might be taken as evidence for the potential efficacy of a drug affecting the risk factor pathway. For the reasons given above, however, absence of evidence for such an association does not necessarily imply lack of efficacy. Although we can 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 might 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 relations between modifiable risk factors and outcomes of interest. It is one of the few epidemiological methods that can help in the selection of targets for therapeutic intervention. It would be misleading, however, 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 might not correspond to the magnitude of the effect in practice.

Contributors: SB was the lead author; the other authors were involved with all stages in the conceptualising and editing of this article. SB is guarantor.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests: All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://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 currently on secondment to the University of Cambridge from Pfizer Pharmaceuticals.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 1 Lawlor D, Harbord R, Sterne J, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27:1133-63. doi:10.1002/sim.3034
- 2 Kamstrup P, Tybjaerg-Hansen A, Steffensen R, Nordestgaard B. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA* 2009;301:2331-9. doi:10.1001/jama.2009.801
- 3 CRP CHD Genetics Collaboration. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ* 2011;342:d548. doi:10.1136/bmj.d548
- 4 Davey Smith G, Ebrahim S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* 2005;330:1076-9. doi:10.1136/bmj.330.7499.1076
- 5 Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007;16:309-30. doi:10.1177/0962280206077743
- 6 Waterworth D, Ricketts S, Song K, Chen L, Zhao JH, Ripatti S, et al. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* 2010;30:2264-76. doi:10.1161/atvbaha.109.201020
- 7 Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1-22. doi:10.1093/ije/dyg070
- 8 Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;33:30-42. doi:10.1093/ije/dyh132
- 9 Nitsch D, Molokhia M, Smeeth L, DeStavola B, Whittaker J, Leon D. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *Am J Epidemiol* 2006;163:397-403. doi:10.1093/aje/kwj062
- 10 Dekkers O, von Elm E, Algra A, Romijn J, Vandenbroucke J. How to assess the external validity of therapeutic trials: a conceptual approach. *Int J Epidemiol* 2010;39:89-94. doi:10.1093/ije/dyp174
- 11 Tanabe M, Matsumoto T, Shibuya K, Tateda K, Miyazaki S, Nakane A, et al. Compensatory response of IL-1 gene knockout mice after pulmonary infection with *Klebsiella pneumoniae*. *J Med Microbiol* 2005;54:7-13. doi:10.1099/jmm.0.45736-0

### 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 that is associated with the risk factor divides the population into groups that 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 can be misleading
- Overinterpreting a null finding: The differences in the risk factor between the genetic groups are usually small compared with its overall variation. A null finding might 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. In some cases sample sizes in tens of thousands are required to provide sufficient power to reliably interpret a null finding<sup>9</sup>
- Overinterpreting a positive finding: While the Mendelian randomisation hypothesis relates to genetic groups, one aim of a Mendelian randomisation is to determine 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 might not directly translate into the observed effect on the outcome of modifying the risk factor in practice.

### Summary points

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

The effects of such interventions can be quantitatively different to those obtained from Mendelian randomisation

- Hirschfield G, Gallimore J, Kahan M, Hutchinson WL, Sabin CA, Benson GM, et al. Transgenic human C-reactive protein is not proatherogenic in apolipoprotein E-deficient mice. *Proc Natl Acad Sci U S A* 2005;102:8309-14. doi:10.1073/pnas.0503202102
- Brennan P, McKay J, Moore L, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, et al. Obesity and cancer: Mendelian randomization approach utilizing the *FTO* genotype. *Int J Epidemiol* 2009;38:971-5. doi:10.1093/ije/dyp162
- Wardle J, Carnell S, Haworth C, Farooqi I, O'Rahilly S, Plomin R. Obesity associated genetic variation in *FTO* is associated with diminished satiety. *J Clin Endocrinol Metab* 2008;93:3640-3. doi:10.1210/jc.2008-0472
- Cheung B, Lauder I, Lau C, Kumana C. Meta-analysis of large randomized controlled trials to evaluate the impact of statins on cardiovascular outcomes. *Br J Clin Pharmacol* 2004;57:640-51. doi:10.1111/j.1365-2125.2003.02060.x
- Taylor F, Ward K, Moore T, Burke M, Davey Smith G, Casas JP, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev* 2011;1:CD004816.
- Law M, Wald N, Rudnicka A. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. *BMJ* 2003;326:1423. doi:10.1136/bmj.326.7404.1423
- International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011;478:103-9. doi:10.1038/nature10405
- Law M, Morris J, Wald N. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *BMJ* 2009;338:b1665. doi:10.1136/bmj.b1665

Accepted: 21 September 2012

Cite this as: *BMJ* 2012;345:e7325

© BMJ Publishing Group Ltd 2012

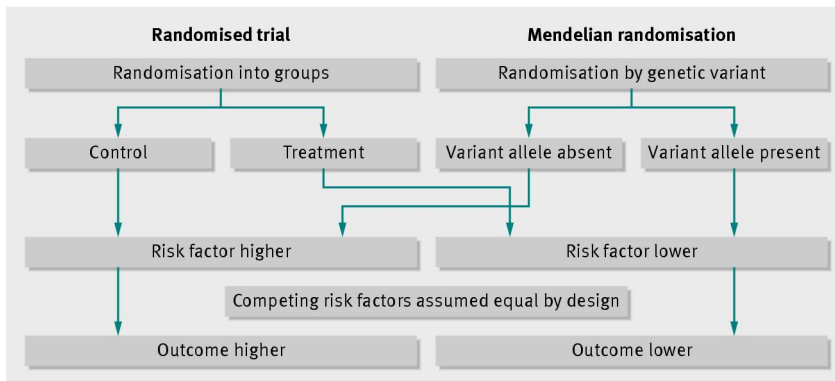
## Table

**Table 1 | Association of five genetic variants with log transformed concentrations of low density lipoprotein cholesterol and risk of coronary heart disease taken from Waterworth et al.<sup>6</sup> Causal estimates (with 95% confidence intervals) of odds ratio (OR) for 30% reduction in low density lipoprotein cholesterol on coronary heart disease from Mendelian randomisation with each genetic variant in turn**

Genetic variant (relevant gene)	Per allele change in log transformed low density lipoprotein cholesterol (SE)	Per allele OR of coronary heart disease (95% CI)	OR of coronary heart disease per 30% decrease in low density lipoprotein cholesterol*
rs11206510 (PCSK9)	-0.026 (0.004)	0.93 (0.88 to 0.99)	0.40 (0.15 to 0.85)
rs660240 (SORT1)	-0.044 (0.004)	0.85 (0.80 to 0.90)	0.27 (0.15 to 0.44)
rs515135 (APOB)	-0.038 (0.004)	0.90 (0.85 to 0.96)	0.37 (0.19 to 0.66)
rs12916 (HMGCR)	-0.023 (0.003)	0.94 (0.90 to 0.99)	0.38 (0.16 to 0.80)
rs2738459 (LDLR)	-0.018 (0.004)	0.96 (0.89 to 1.03)	0.45 (0.07 to 1.95)

\*30% decrease in concentration is equivalent to change in log transformed concentration of -0.357.

## Figure



Comparison of Mendelian randomisation and randomised trial