Expressing Results From a Mendelian Randomization Analysis: Separating Results From Inferences

Stephen Burgess, PhD,
Medical Research Council Biostatistics Unit, Cambridge Institute of Public Health, University of Cambridge, Cambridge, United Kingdom; Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

Christopher J. O'Donnell, MD,MPH,
Cardiology Section, Veteran's Administration Boston Healthcare System, Boston, Massachusetts; Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

Dipender Gill, BMBCh, PhD
Department of Epidemiology and Biostatistics, Medical School Building, St Mary's Hospital, Imperial College London, London, United Kingdom; Clinical Pharmacology and Therapeutics Section, St George's, University of London, London, United Kingdom

Choice of language is important, particularly relatingto causal claims. Mendelian randomization is an epidemiologic approach used in biomedical research to assess the evidence for a causal hypothesis using genetic associations estimated in observational data. Increasingly, mendelian randomization investigations are providing evidence supporting the potential effectiveness of treatments for which causality has not been established in a randomized clinical trial. The language used to describe findings from mendelian randomization investigations is inconsistent, with some studies claimingthat mendelian randomization can demonstrate that an exposure has a causal effect on an outcome and others making more circumspect claims. As an example, 2 articles investigating height and coronary artery disease (CAD) using mendelian randomization were published in 2015. One article reported on the "causal effect of completed growth, measured by adult height, on coronary heart disease,"[1] while the other article reported "a genetic approach to investigate the association between height and CAD."[2] Herein, we explore the assumptions of mendelian randomization and discuss howto interpret and express results findings from such an analysis.

In basic terms, a mendelian randomization investigation takes genetic variants associated with a modifiable exposure and assesses whether these same variants are associated with an outcome. If the genetic variants satisfy the assumptions of an instrumental variable, then

a potential causal association of the exposure on the outcome may be inferred from an association of these variants with the outcome.[3] The instrumental variable assumptions state that a genetic variant influences the distribution of the exposure in the population, but it is not associated with competing risk factors on alternative pathways to the outcome and it does not influence the outcome directly. This implies that a genetic variant acts similarly to randomized treatment allocation in a randomized clinical trial, defining subgroups of the population that differ systematically with respect to the exposure of interest but not with respect to other factors.[4]

While Mendel's laws of inheritance, choice of sexual partner, and the fixed nature of the genome all provide general plausibility to the use of genetic variants as instrumental variables, the instrumental variable assumptions can never be empirically validated for any particular genetic variant. Genetic variants used to proxy the potential effect of intervening on an exposure may have pleiotropic associations that affect the outcome through pathways unrelated to the exposure. Additionally, genetic associations may reflect differences in allele frequencies between strata of the population (such as different ethnic groups) rather than biological mechanisms. Such violations of the instrumental variable assumptions can bias mendelian randomization estimates.

We encourage researchers to separate the factual description of analysis results from any inference that is made. Mendelian randomization assesses whether genetic predictors of an exposure are associated with the outcome or not. Equivalently, it assesses whether genetically predicted values of the exposure are associated with the outcome or not. This is a plain statement of the analysis results and does not rely on any assumption or make any causal claim. As an aside, we prefer the term *genetically predicted* over *genetically determined* because relationships between genetic variants and risk factors are rarely deterministic.

The inference that is often made from a mendelian randomization analysis is that intervention on or change in the exposure would lead to a change in the outcome. This is a causal claim and relies on an untestable assumption. Specifically, we are assuming that differences in the outcome arising from the effect of a genetic variant on the exposure (which often represent differences between genetic subgroups in the trajectory of the exposure across the life course[5]) are informative about what would happen if we intervened on the exposure directly. As we discuss below, there are often substantive differences between genetic variants that are proxies for an exposure and clinical interventions for the exposure,[6] which can lead to quantitative differences in estimates. If the instrumental variable assumptions are satisfied, then the presence and direction of the genetic association with the outcome is informative of the presence and direction of the association between the outcome and intervention for the exposure in practice. If the instrumental variable assumptions are in doubt, then a weaker conclusion of potentially shared genetic predictors may be more reasonable,[7] particularly if genetic associations with the outcome are inconsistent across different variants.

We recommend that mendelian randomization findings be expressed in 2 parts: first, the factual statement that genetically predicted values of the exposure are associated with

the outcome, and second (if appropriate), any claim relating to a causal inference that an exposure is a potentially causal determinant of the outcome. The latter should be accompanied with appropriate caution regarding limitations of mendelian randomization analyses. The degree of confidence in a potentially causal conclusion should depend on the plausibility of the instrumental variable assumptions being satisfied and a broad assessment of the quality of the study.

As for numerical mendelian randomization estimates, these relate primarily to the magnitude of association between genetically predicted levels of the exposure and the outcome. Under the instrumental variable assumptions, a mendelian randomization estimate has a potential for causal inference similar to changing one's genotype from conception. However, this is typically not the target parameter of clinical interest. Given the many differences between genetic and clinical interventions in terms of timing, duration, scale, and mechanism, mendelian randomization estimates are likely to differ from causal associations that are seen in practice.[8] In particular, mendelian randomization estimates tend to be larger in magnitude per unit change in the exposure because they represent lifelong rather than short-term effects. So while it is appropriate to provide an estimate in terms of the association with the outcome per unit change in genetically predicted values of the exposure, such estimates should not be thought of as the predicted real-world influence of changes to the exposure.

For example, while higher genetically predicted levels of lipoprotein(a) have been associated with increased CAD risk,[9] any potential clinical effect of lowering lipoprotein(a) is likely to be lesser in magnitude than the estimate from a mendelian randomization analysis, as has previously been observed for low-density lipoprotein cholesterol.[6] Although this mendelian randomization analysis provides evidence supporting a causal inference that lipoprotein(a) detrimentally affects CAD risk, it does not directly demonstrate causation. Similarly, genetically predicted blood pressure levels have been shown to be associated with valvular heart disease. The potential causal inference from this observation is that lifelong elevated blood pressure levels increase valvular heart disease risk.[10]

In summary, a causal inference from a mendelian randomization analysis relies on the assumption that the selected genetic variants are appropriate for use as instrumental variables, that is, they are unconfounded by other unobserved variables or biases. Findings of mendelian randomization analyses, and in particular numerical estimates, should primarily be presented in terms of the presence and magnitude of the association between genetically predicted levels of the exposure and the outcome. Any statement regarding a causal hypothesis is a subjective inference that the analysts have made and not something that has been demonstrated directly from the data. Therefore, statements about potential causal inferences should be presented separately and secondarily to the primary factual report of the analysis result.

References

1. Nüesch E, Dale C, Palmer TM, et al. EPIC-Netherland Investigators; UCLEB Investigators; IN Day. Adult height, coronary heart disease and stroke: a multi-locus Mendelian randomization meta-analysis. Int J Epidemiol. 2016; 45 (6) 1927–1937. DOI: 10.1093/ije/dyv074 [PubMed: 25979724]

2. Nelson CP, Hamby SE, Saleheen D, et al. CARDIoGRAM+C4D Consortium. Genetically determined height and coronary artery disease. N Engl J Med. 2015; 372 (17) 1608–1618. DOI: 10.1056/NEJMoa1404881 [PubMed: 25853659]

3. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res. 2007; 16 (4) 309–330. DOI: 10.1177/0962280206077743 [PubMed: 17715159]

4. Thanassoulis G, O'Donnell CJ. Mendelian randomization: nature's randomized trial in the post-genome era. JAMA. 2009; 301 (22) 2386–2388. DOI: 10.1001/jama.2009.812 [PubMed: 19509388]

5. Labrecque JA, Swanson SA. Interpretation and potential biases of Mendelian randomization estimates with time-varying exposures. Am J Epidemiol. 2019; 188 (1) 231–238. DOI: 10.1093/aje/kwy204 [PubMed: 30239571]

6. Ference BA, Yoo W, Alesh I, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. J Am Coll Cardiol. 2012; 60 (25) 2631–2639. DOI: 10.1016/j.jacc.2012.09.017 [PubMed: 23083789]

7. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. J Clin Epidemiol. 2016; 69: 208–216. DOI: 10.1016/j.jclinepi.2015.08.001 [PubMed: 26291580]

8. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. BMJ. 2012; 345 e7325 doi: 10.1136/bmj.e7325 [PubMed: 23131671]

9. Burgess S, Ference BA, Staley JR, et al. European Prospective Investigation Into Cancer and Nutrition–Cardiovascular Disease (EPIC-CVD) Consortium. Association of LPA variants with risk of coronary disease and the implications for lipoprotein(a)-lowering therapies: a Mendelian randomization analysis. JAMA Cardiol. 2018; 3 (7) 619–627. DOI: 10.1001/jamacardio.2018.1470 [PubMed: 29926099]

10. Nazarzadeh M, Pinho-Gomes AC, Smith Byrne K, et al. Systolic blood pressure and risk of valvular heart disease: a Mendelian randomization study. JAMA Cardiol. 2019; 4 (8) 788–795. DOI: 10.1001/jamacardio.2019.2202 [PubMed: 31290937]