ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors

Stephen Burgess[a,b,*], Adam S. Butterworth[a], John R. Thompson[c]

[a]*Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, 2 Worts Causeway, Cambridge CB1 8RN, UK*
[b]*Homerton College, University of Cambridge, Hills Road, Cambridge CB2 8PH, UK*
[c]*Department of Health Sciences, Adrian Building, University Road, Leicester, LE1 7RH, UK*

Accepted 7 August 2015; Published online xxxx

Abstract

Objective: Mendelian randomization is a popular technique for assessing and estimating the causal effects of risk factors. If genetic variants which are instrumental variables for a risk factor are shown to be additionally associated with a disease outcome, then the risk factor is a cause of the disease. However, in many cases, the instrumental variable assumptions are not plausible, or are in doubt. In this paper, we provide a theoretical classification of scenarios in which a causal conclusion is justified or not justified, and discuss the interpretation of causal effect estimates.

Results: A list of guidelines based on the 'Bradford Hill criteria' for judging the plausibility of a causal finding from an applied Mendelian randomization study is provided. We also give a framework for performing and interpreting investigations performed in the style of Mendelian randomization, but where the choice of genetic variants is statistically, rather than biologically motivated. Such analyses should not be assigned the same evidential weight as a Mendelian randomization investigation.

Conclusion: We discuss the role of such investigations (in the style of Mendelian randomization), and what they add to our understanding of potential causal mechanisms. If the genetic variants are selected solely according to statistical criteria, and the biological roles of genetic variants are not investigated, this may be little more than what can be learned from a well-designed classical observational study.

Keywords: Mendelian randomization; Instrumental variable; Causal inference; Genetic variants; Genetic predictors; Aetiology; Translational Genetics

1. Introduction

Genome-wide association studies have revealed genetic predictors of many clinically relevant traits, including modifiable risk factors and disease outcomes. Many investigators have taken two such traits and considered the statistical question of whether genetic variants that are associated with trait A (often taken to be a risk factor and viewed as a putative cause) also show an association with trait B (often taken to be a disease outcome), for example, under the heading of Mendelian randomization [1,2]. However, conclusions from such analyses have been diverse, ranging from a direct causal interpretation (trait A causes trait B) to one of shared etiology (trait A and trait B have common predictors). In this article, we consider conditions under which a causal interpretation is justified and discuss situations in which weaker conclusions are more appropriate.

2. Classification of scenarios

We consider the following classification of possible scenarios for the relationship between two variables A and B such that genetic variant(s) associated with A are also associated with B. An interventional definition of causality is presumed; A is a cause of B means that intervention on the distribution of A results in changes to the distribution of B [3]. We assume that either logic or biological knowledge is able to provide an ordering between A and B by which A is the putative cause and B is the putative effect. The three scenarios we consider are as follows:

Conflict of Interest/Financial Disclosure: The authors have no conflict of interest regarding this article. The authors have no relevant financial interests to disclose regarding this article.

Funding: S.B. is supported by the Wellcome Trust (grant number 100114).

* Corresponding author. Department of Public Health and Primary Care, Strangeways Research Laboratory, 2 Worts Causeway, Cambridge CB1 8RN, UK. Tel.: +44 1223 748651; fax: +44 1223 748658.

E-mail address: sb452@medschl.cam.ac.uk (S. Burgess).

http://dx.doi.org/10.1016/j.jclinepi.2015.08.001

What is new?

- The popularity of Mendelian randomization as a tool for investigating causal relationships in observational data is currently increasing.
- There is a distinction between Mendelian randomization as it was initially conceived and performed (mainly for circulating biomarkers using few genetic variants in relevant gene regions) and how it is often used today (often opportunistically using large numbers of genetic variants whose functional relevance is unknown).
- Biological guidelines based on the Bradford Hill criteria, and statistical criteria based on associations with measured covariates and homogeneity of evidence across genetic variants are given to help judge the plausibility of a causal conclusion.
- Causal claims should be reserved to cases where the evidence for the instrumental variable assumptions is strong; otherwise the language of common genetic predictors should be used.

1. A is a cause of B, and all causal pathways from the genetic variant(s) to B pass through A;
2. A is a cause of B, but there are alternative causal pathways leading from the genetic variant(s) to B which do not pass through A;
3. A is not a cause of B—the genetic variant(s) are independently associated with A and B.

Diagrams representing the relationships between the variables in each case are given in Fig. 1. We continue to explore each of the scenarios mentioned previously in turn.

2.1. All causal pathways through risk factor

To infer a causal effect of A on B, it is necessary that genetic variants used in the analysis satisfy the assumptions of an instrumental variable [4,5]:

1. The set of genetic variants is associated with the risk factor A;
2. Each genetic variant is independent of confounders of the association between A and B;
3. If the risk factors were kept constant, intervention on the genetic variant(s) would not have an effect on the outcome.

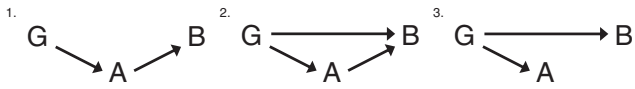

Fig. 1. Diagrams illustrating scenarios of causal relationships between selected genetic variant(s) G, putative causal trait A, and putative effect trait B, compatible with genetic variant(s) being associated with both traits.

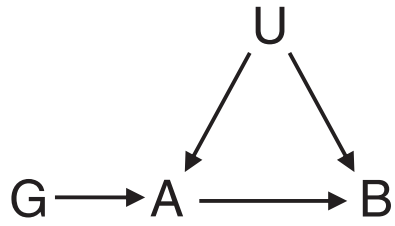

Fig. 2. Diagram illustrating causal relationships between genetic variant(s) G, putative causal trait (risk factor) A, putative effect trait (outcome) B, and confounders U necessary for instrumental variable assumptions to be satisfied.

A diagram corresponding to these assumptions is given in Fig. 2.

These assumptions imply that all causal pathways from the genetic variant(s) to the outcome pass through the putative causal risk factor, and there are no alternative pathways not via the risk factor [6]. Formal considerations about how causal pathways are defined are given in the Appendix at www.jclinepi.com. The assumptions require that genetic variants used for the assessment of the causal nature of a risk factor must be specific in their associations with the risk factor, although they may also show associations with other variables via downstream effects of the risk factor. For example, genetic variants that are candidate instrumental variables for body mass index (BMI) may show associations with C-reactive protein (CRP), because of a causal effect of BMI on CRP [7]. This means that the genetic variants can have associations with other variables via mediation (ie, the genetic association with the other variable is mediated via the risk factor of interest), but not via pleiotropy (ie, the genetic association with the other variable is via a different causal pathway and not via the risk factor of interest; Fig. 3).

If we seek to assess whether there is a causal effect of A on B, but not to provide an estimate of a causal effect parameter, then only the three instrumental variable assumptions listed previously are required. Under these assumptions, an association between the outcome B and genetic variants which are instrumental variables for A implies a causal effect of A on B [8]. To estimate a causal effect parameter, further assumptions are required [9], including linearity of the risk factor–outcome association, and the stable unit treatment value assumption (the value of the outcome for each individual depends on the value of the

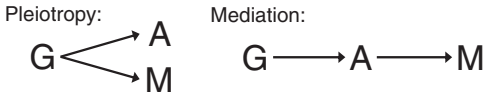

Fig. 3. Diagrams illustrating the difference between pleiotropy (left), where genetic variant G is independently associated with traits A and M, and mediation (right), where G is associated with trait M only via the effect of A.