Mendelian randomisation studies of type 2 diabetes: future prospects

Manjinder S Sandhu[1,2,‡], **Sally L Debenham**[1,2], **Inês Barroso**[3], and **Ruth JF Loos**[1]

[1] MRC Epidemiology Unit, Strangeways Research Laboratory, Cambridge, UK

[2] Department of Public Health & Primary Care, Institute of Public Health, University of Cambridge, UK

[3] Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

The integration of principles and procedures underlying population genetics and epidemiology provides a potential research framework to make causal inferences about the association between a risk factor and disease (1;2). In an epidemiological context, numerous biomarkers have been linked to the development of type 2 diabetes. In this issue of Diabetologia, Herder and colleagues assess the prospective association between circulating levels of macrophage inhibitory factor and risk of type 2 diabetes (3). Their results provide suggestive statistical evidence for a sex-specific association between this biomarker and risk of type 2 diabetes. This observation is consistent with the current paradigm that inflammation plays a role in the pathogenesis of insulin resistance and type 2 diabetes (4).

The association might be causal. Alternatively, given the observational nature of epidemiological research, it may also be completely explained by confounding. In this scenario, the relation between the biomarker and disease is explained by a third factor that is associated with the biomarker and disease risk. That is, the biomarker is not independently or causally linked to disease risk. Generally, epidemiological approaches use multivariable statistical analysis to reduce confounding. However, because of residual confounding (as a result of measurement error) and unknown confounders, standard observational epidemiology cannot resolve whether an observed association has a causal basis. The conceptual ambiguity of any disease mechanism (specifically, defining mediators and confounders) also limits statistical modelling to reduce confounding. Furthermore, the observed association may be due to reverse association/causation—that is, the relation between biomarker and disease could be the result of undiagnosed or early disease rather than the risk factor, and thus a consequence of disease rather than a cause.

'Mendelian randomisation'—the random assortment of genes from parents to offspring that occurs during gamete formation and conception (1)—provides an epidemiological approach that is much less susceptible to confounding by classical or environmental risk factors and excludes reverse causality as a possible non causal explanation for the observed association between a biomarker and disease risk (5). In this research framework, the interrelation and consistency of associations among genetic variants that encode or regulate a biomarker, the expression or circulating levels of the biomarker, and disease risk may characterise the true magnitude of the relation between the biomarker and risk of disease (Figure 1) (5-7). Reproducible evidence for statistical association between a genetic variant and biomarker provide the basis for a Mendelian randomisation study. Because of independent assortment (1), characteristics that may confound any association between a biomarker and disease are

‡ Correspondence should be addressed to Manjinder S Sandhu (manj.sandhu@srl.cam.ac.uk).

equally distributed among the relevant genetic variants. Thus a comparison of groups of individuals defined by a genetic variant, based on a Mendelian randomisation design, is equivalent to a randomised comparison, with only the relevant biomarker differing across the relevant genetic variant (8;9). Examining the relation between variants in the genome that show unequivocal associations with the relevant biomarker and disease risk, is, therefore, a potential method of assessing whether a biomarker might be causally linked to disease. In this context, Herder and colleagues, based on specificity and directional consistency of associations among genetic variants, biomarker and disease risk, provide some evidence for a possible causal link between circulating levels of macrophage inhibitory factor and risk of type 2 diabetes (3). However, much more robust statistical evidence is required to make causal inferences using this approach (6;10-12).

Mendelian randomisation studies, like other genetic epidemiological studies, require reliable identification of statistical associations between genetic variants and biomarkers and disease risk (12). Importantly, the principal requirement is the ability of Mendelian randomisation studies to detect an association of equivalent magnitude to that predicted by a proportional change in the relevant trait or risk factor. Given that individual genetic variants are likely to only explain a very small proportion of the variation in a biomarker trait or risk factor, these assessments will necessitate very large sample sizes, which are probably beyond the scale of existing studies or collaborations in type 2 diabetes genetics (13). This statistical limitation is compounded by random measurement error in the assessment of the magnitude of the associations among the genetic variant, biomarker and disease risk, which can lead to attenuated association signals. To help overcome issues of statistical resolution, investigators have opted to use meta-analysis (8). However, using this approach for Mendelian randomisation triangulation has several limitations, which are generic to this strategy (14). Heterogeneity among studies (for example, due to potential gene–gene or gene–environment interactions) may distort assessments of the magnitude of the associations (6). This potential limitation may be exacerbated when different studies are used for data aggregation for each side of the Mendelian randomisation triangle (Figure 1).

It is, therefore, likely that international collaborative frameworks for Mendelian randomisation studies of type 2 diabetes will be needed to achieve the required level of statistical precision. For example, the magnitude of reproducible statistical associations between genetic variants and risk of type 2 diabetes are in the order of around 10-30% increase risk per allele (13). Under a log additive model, around 15,000 cases and 15,000 controls are required to detect a per allele risk of 10% (with an alpha of 1×10^{-4} and 90% power) for a relatively common genetic variant (20% allele frequency). The association between genetic variants and type 2 diabetes risk that also show reproducible associations with relevant biomarkers may be even smaller (15). Thus it is likely that Mendelian randomisation studies will need much greater statistical resolution. To facilitate this requirement, improved statistical approaches for meta-analysing data across studies are in development (16). However, data pooling strategies have distinct advantages in this context, allowing enhanced data harmonisation and statistical testing, including haplotype analysis and imputational methods to increase comparability of markers across studies.

There are other caveats. Genetic confounding and biological compensation ("canalization") may also limit inferences in Mendelian randomisation studies. Biological compensation (a developmental and physiological adaptation to a genetic difference) could lead to inconsistencies in the magnitude of the triangulated associations (17). This physiological adaptation, among other important differences, distinguishes Mendelian randomisation studies from randomised controlled trials, which are not susceptible to this limitation (6). Importantly, the pleiotropic nature (multiple biological effects) of some genetic variants may produce confounded associations between genetic variants and phenotypes. Undetected

population stratification (confounding), where genetic differences and traits associations may be due to underlying differences in ancestry, may also distort the magnitude of gene–biomarker and gene–disease associations. By contrast, correlation among genetic variants (linkage disequilibrium) can be used to examine associations among these variants, traits and diseases without directly ascertaining functional variants. However, in the same context, correlation with pleiotropic genetic variants, for example, could distort associations.

Despite these limitations, Mendelian randomisation provides a potential research framework to assess causal links between biological and environmental phenotypes and disease risk. These studies, when correctly performed, will provide insights into aetiological mechanisms and causality, informing potential therapeutic and preventative strategies.

Acknowledgments

References

1. Davey SG, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003; 32:1–22. [PubMed: 12689998]

2. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet. 2001; 358:1356–1360. [PubMed: 11684236]

3. Herder C, Klopp N, Baumert J, Müller M, Khuseyinova N, Meisinger C, Martin S, Illig T, Koenig W, Thorand B. Effect of Macrophage Migration Inhibitory Factor (MIF) Gene Variants and MIF Serum Concentrations on the Risk of Type 2 Diabetes: Results From the MONICA/KORA Augsburg Case-Cohort Study, 1984-2002. 2007 Diabetologia In Press.

4. Wellen KE, Hotamisligil GS. Inflammation, stress, and diabetes. J Clin Invest. 2005; 115:1111–1119. [PubMed: 15864338]

5. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey SG. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med. 2007

6. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. Am J Epidemiol. 2006; 163:397–403. [PubMed: 16410347]

7. Hingorani A, Humphries S. Nature's randomised trials. Lancet. 2005; 366:1906–1908. [PubMed: 16325682]

8. Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: evidence on a causal link from mendelian randomisation. Lancet. 2005; 365:224–232. [PubMed: 15652605]

9. Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day IN, Palmer LJ, Hattersley AT, Ebrahim S, Lowe GD, Rumley A, Davey SG. C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. Lancet. 2005; 366:1954–1959. [PubMed: 16325697]

10. Minelli C, Thompson JR, Tobin MD, Abrams KR. An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization. Am J Epidemiol. 2004; 160:445–452. [PubMed: 15321841]

11. Tobin MD, Minelli C, Burton PR, Thompson JR. Commentary: development of Mendelian randomization: from hypothesis test to 'Mendelian deconfounding'. Int J Epidemiol. 2004; 33:26–29. [PubMed: 15075142]

12. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr. Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS. Replicating genotype-phenotype associations. Nature. 2007; 447:655–660. [PubMed: 17554299]

13. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat Rev Genet. 2007; 8:657–662. [PubMed: 17703236]

14. Egger M, Schneider M, Davey SG. Spurious precision? Meta-analysis of observational studies. BMJ. 1998; 316:140–144. [PubMed: 9462324]

15. Casas JP, Shah T, Cooper J, Hawe E, McMahon AD, Gaffney D, Packard CJ, O'Reilly DS, Juhan-Vague I, Yudkin JS, Tremoli E, Margaglione M, Di Minno G, Hamsten A, Kooistra T, Stephens JW, Hurel SJ, Livingstone S, Colhoun HM, Miller GJ, Bautista LE, Meade T, Sattar N, Humphries SE, Hingorani AD. Insight into the nature of the CRP-coronary event association using Mendelian randomization. Int J Epidemiol. 2006; 35:922–931. [PubMed: 16565153]

16. Higgins JP, Little J, Ioannidis JP, Bray MS, Manolio TA, Smeeth L, Sterne JA, Anagnostelis B, Butterworth AS, Danesh J, Dezateux C, Gallacher JE, Gwinn M, Lewis SJ, Minelli C, Pharoah PD, Salanti G, Sanderson S, Smith LA, Taioli E, Thompson JR, Thompson SG, Walker N, Zimmern RL, Khoury MJ. Turning the pump handle: evolving methods for integrating the evidence on gene-disease association. Am J Epidemiol. 2007; 166:863–866. [PubMed: 17804859]

17. Davey SG, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ. 2005; 330:1076–1079. [PubMed: 15879400]
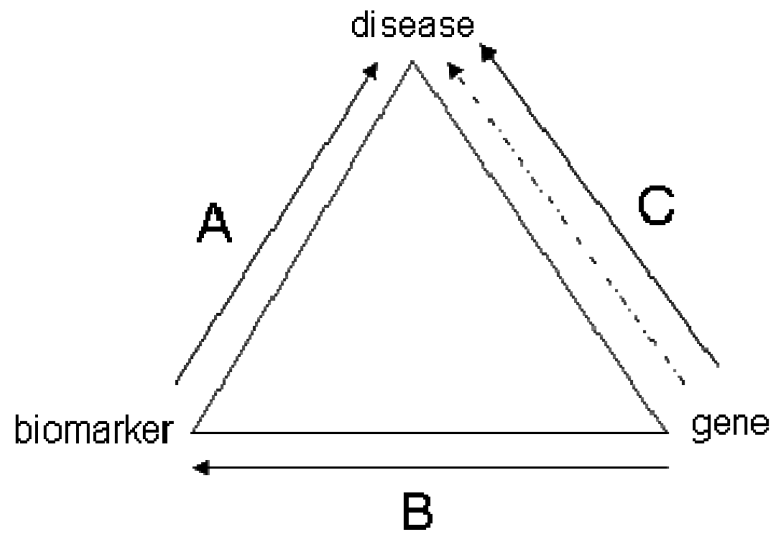
Figure 1.
Mendelian triangulation: strategy to help clarify whether there is a causal relation between a biomarker and disease risk. Information on the magnitude of the association between the biomarker and disease (A), combined with information on the magnitude of the association between the genetic variant and the biomarker (B) is used to estimate the expected magnitude of the association between the genetic variant and disease risk (dashed line - C). Based on several assumptions (see text), including a linear association between the biomarker and disease, the direct assessment of C provides an unconfounded assessment of the association between the biomarker and disease risk (solid line – C).