Published by Oxford University Press on behalf of the International Epidemiological Association

Letters to the Editor

A cautionary note on the use of Mendelian randomization to infer causation in observational epidemiology

From MURIELLE BOCHUD,[1]* ARNAUD CHIOLERO,[1] ROBERT C ELSTON[2] and FRED PACCAUD[1]

The concept of Mendelian randomization when used in the context of association studies refers to the random allocation of alleles at the time of gamete formation. This concept has received a lot of attention in recent years in observational epidemiology for its potential to reduce residual confounding and to protect against reverse causation.[1] In brief, the random segregation of alleles at the time of gamete formation is a natural experiment that occurs before the outcome of interest, which protects against falsely reversing the inference about a cause-effect relationship. If a particular genetic variant is strongly associated with a risk factor for the outcome of interest, the association between this genetic variant and the outcome of interest may be used to infer causality.

Table 1 indicates the commonly acknowledged necessary conditions (1–9) for Mendelian randomization to provide this protection in observational epidemiology.[2,3] We list additional necessary conditions (10–12) that have been given little attention and may be of relevance for certain genetic variants.

The first of these additional conditions is the absence of *transmission ratio distortion* (TRD). TRD occurs when the distribution of alleles at a particular locus differs in the surviving offspring from that expected on the basis of Mendelian proportions. TRD can occur either during or after meiosis. *Segregation distortion* or *meiotic drive* refers to the phenomenon of non-random assignment of alleles at the time of gamete formation and therefore describes TRD during meiosis.[4] After meiosis, TRD may result from *selective survival* between conception and birth or later. In humans, a few examples of TRD have been reported.[5–10] In particular, a recent report suggests that transmission distortion may be widespread in the human genome.[6] Selective survival due to the genetic variant of interest is acknowledged as a potential source of bias analogous to differential loss of follow up in randomized controlled trials.[2] With that perspective, selective survival may occur not only between conception and birth, but also between birth and entry into the study if the genetic variant of interest causes early mortality.

The second additional condition is the absence of *parent-of-origin effect*. A parent-of-origin effect occurs whenever the effect of an allele on a phenotype of interest depends on whether the allele was passed on to the child from the father or from the mother. Some genes may be silent or functionally active depending on whether a particular variant was maternally or paternally inherited. This effect is mediated through an epigenetic modification called imprinting. As a consequence, a parent-of-origin effect implies that the effect on the phenotype conferred by a specific genetic variant is not homogeneous in the population.

Segregation distortion and parent-of-origin effect have, to our knowledge, not been previously described as potential limitations to the use of Mendelian randomization for inferring causality in observational epidemiology. We consider here data on the relationship between homocysteine blood level and stroke to illustrate how these limitations may jeopardize the use of Mendelian randomization to infer causation.

Homocysteine blood level is associated with increased risk of stroke[11] and is determined, in part, by the activity of methylenetetrahydrofolate reductase (MTHFR). The *C677T MTHFR* variant is associated with both the risk of stroke and homocysteine blood level.[12] The risk of stroke conferred by the *C677T MTHFR* variant is consistent with the risk expected from the increased homocysteine blood level associated with the *C677T MTHFR* variant.[12] These findings strongly suggest a causal relationship between homocysteine and stroke. We now consider to what extent the earlier-mentioned conditions may not have been met for this particular example.

Infante-Rivard and Weinberg[13] assessed TRD for variants in five genes associated with thrombophilic disorders, including the *MTHFR* gene. They found some evidence of TRD for the *MTHFR* gene, which may be due to either segregation distortion at the time of gamete formation or due to post-meiotic events occurring between conception and birth.

[1] Community Prevention Unit, University Institute of Social and Preventive Medicine, Lausanne, Switzerland.

[2] Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, USA.

* Corresponding author. Community Prevention Unit, University Institute of Social and Preventive Medicine, Lausanne, Switzerland.
E-mail: murielle.bochud@chuv.ch

Table 1 Necessary conditions for the use of Mendelian randomization to infer causality in observational epidemiology

1	There are enough data to establish reliable genotype-intermediate phenotype, or genotype-trait, associations
2	There is no confounding due to linkage disequilibrium
3	There is no confounding due to population stratification
4	There is no pleiotropy
5	There is no canalization nor developmental compensation (i.e. a functional adaptation to a specific genotype influencing the expected genotype-disease association)
6	A suitable genetic variant exists to study the exposure of interest
7	The association between gene and gene product is strong
8	The effects of a gene on a disease outcome acts only via the intermediate phenotype
9	The genetically determined exposure has a similar impact on the disease outcome as the environmental exposure investigated
10	There is no segregation distortion at the locus of interest
11	There is no selective survival due to the genetic variant of interest
12	There is no parent-of-origin effect

Adapted and expanded from Nitsch *et al*.[2] and Davey Smith and Ebrahim.[3]

Furthermore, fetal *C677T MTHFR* status has been associated with fetal viability, i.e. selective survival between gamete formation and birth.[14,15] Maternal *C677T MTHFR* status may influence pregnancy outcome. The *C677T* status of the mother has been associated with idiopathic recurrent pregnancy losses in some studies[16–18] but not in others.[19–22] Maternal hyperhomocysteinemia could be a risk factor for recurrent early pregnancy loss.[23] Together, these observations may contribute to TRD at this particular locus and so invalidate the assumption of random segregation of alleles that underlies Mendelian randomization. Finally, although there is currently no evidence for a parent-of-origin effect, few studies have evaluated genomic imprinting at the *C677T MTHFR* locus[13] so far. Due to the growing evidence for the role of imprinted genes in common complex diseases,[24–28] assuming that parent-of-origin effects are absent may be disputable in many instances.

Hence, for this specific example, not only may the process of randomization itself be flawed, but substantial selective survival may also occur between the time of randomization (i.e. gamete formation) and the time of birth. As a consequence, the genotype distribution at birth may differ from the distribution at the time of randomization. The influence of maternal *C677T* genotype and maternal homocysteine levels on the fetal genotype distribution may represent additional sources of bias.

Currently, little is known about the mechanisms that distort the distribution of alleles from that expected from Mendel's first law between gamete formation, conception, birth and entry into a study. It is therefore difficult to estimate the extent to which these mechanisms may affect the results of Mendelian randomization studies. However, these mechanisms should not be assumed to be negligible until they have been properly evaluated. Potential for segregation distortion, selective survival and parent-of-origin effect should be kept in mind when conducting Mendelian randomization studies to draw causal inference in observational studies.

References

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

[2] Nitsch D, Molokhia M, Smeeth L *et al*. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *Am J Epidemiol* 2006;**163:**397–403.

[3] Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;**33:**30–42.

[4] Naumova AK, Greenwood CM, Morgan K. Imprinting and deviation from Mendelian transmission ratios. *Genome* 2001;**44:**311–20.

[5] Naumova AK, Leppert M, Barker DF *et al*. Parental origin-dependent, male offspring-specific transmission-ratio distortion at loci on the human X chromosome. *Am J Hum Genet* 1998;**62:**1493–99.

[6] Zöllner S, Wen X, Hanchard NA *et al*. Evidence for extensive transmission distortion in the human genome. *Am J Hum Genet* 2004;**74:**62–72.

[7] Eaves IA, Bennett ST, Forster P *et al*. Transmission ratio distortion at the INS-IGF2 VNTR. *Nat Genet* 1999;**22:** 324–25.

[8] Dean NL, Loredo-Osti JC, Fujiwara TM *et al*. Transmission ratio distortion in the myotonic dystrophy locus in human preimplantation embryos. *Eur J Hum Genet* 2006;**14:**299–306.

[9] Friedrichs F, Brescianini S, Annese V *et al*. Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men. *Hum Genet* 2006;**119:**305–11.

[10] Paterson A, Sun L, Liu XQ. Transmission ratio distortion in families from the Framingham Heart Study. *BMC Genet* 2003;**4:**S48.

[11] Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *Br Med J* 2002;**325:**1202.

[12] Casas JP, Bautista LE, Smeeth L *et al*. Homocysteine and stroke: evidence on a causal link from Mendelian randomisation. *Lancet* 2005;**365:**224–32.

[13] Infante-Rivard C, Weinberg CR. Parent-of-origin transmission of thrombophilic alleles to intrauterine growth-restricted newborns and transmission-ratio distortion in unaffected newborns. *Am J Epidemiol* 2005;**162:**891–97.

[14] Isotalo PA, Wells GA, Donnelly JG. Neonatal and fetal methylenetetrahydrofolate reductase genetic polymorphisms: an examination of C677T and A1298C mutations. *Am J Hum Genet* 2000;**67:**986–90.

[15] Zetterberg H, Regland B, Palmer M *et al*. Increased frequency of combined methylenetetrahydrofolate reductase C677T and A1298C mutated alleles in spontaneously aborted embryos. *Eur J Hum Genet* 2002;**10:** 113–18.

[16] Nelen WL, Steegers EA, Eskes TK *et al*. Genetic risk factor for unexplained recurrent early pregnancy loss. *Lancet* 1997;**350:**861.

[17] Mtiraoui N, Zammiti W, Ghazouani L *et al*. Methylenetetrahydrofolate reductase C677T and A1298C polymorphism and changes in homocysteine concentrations in women with idiopathic recurrent pregnancy losses. *Reproduction* 2006;**131:**395–401.

[18] Unfried G, Griesmacher A, Weismuller W *et al*. The C677T polymorphism of the methylenetetrahydrofolate reductase gene and idiopathic recurrent miscarriage. *Obstet Gynecol* 2002;**99:**614–19.

[19] Parle-McDermott A, Pangilinan F, Mills JL *et al*. A polymorphism in the *MTHFD1* gene increases a mother's risk of having an unexplained second trimester pregnancy loss. *Mol Hum Reprod* 2005;**11:**477–80.

[20] Makino A, Nakanishi T, Sugiura-Ogasawara M *et al*. No association of C677T methylenetetrahydrofolate reductase and an endothelial nitric oxide synthase polymorphism with recurrent pregnancy loss. *Am J Reprod Immunol* 2004;**52:**60–66.

[21] Hohlagschwandtner M, Unfried G, Heinze G *et al*. Combined thrombophilic polymorphisms in women with idiopathic recurrent miscarriage. *Fertil Steril* 2003;**79:**1141–48.

[22] Kobashi G, Kato EH, Morikawa M *et al*. MTHFR C677T Polymorphism and factor V Leiden mutation are not associated with recurrent spontaneous abortion of unexplained etiology in Japanese women. *Semin Thromb Hemost* 2005;**31:**266–71.

[23] Nelen WL, Blom HJ, Steegers EA *et al*. Hyperhomocysteinemia and recurrent early pregnancy loss: a meta-analysis. *Fertil Steril* 2000;**74:**1196–99.

[24] Paterson AD, Naimark DM, Petronis A. The analysis of parental origin of alleles may detect susceptibility loci for complex disorders. *Hum Hered* 1999;**49:**197–204.

[25] Guo YF, Shen H, Liu YJ *et al*. Assessment of genetic linkage and parent-of-origin effects on obesity. *J Clin Endocrinol Metab* 2006;**91:**4001–5.

[26] Dong C, Li WD, Geller F *et al*. Possible genomic imprinting of three human obesity-related genetic loci. *Am J Hum Genet* 2005;**76:**427–37.

[27] Polychronakos C, Kukuvitis A, Giannoukakis N *et al*. Parental imprinting effect at the INS-IGF2 diabetes susceptibility locus. *Diabetologia* 1995;**38:**715–19.

[28] Lindsay RS, Kobes S, Knowler WC *et al*. Genome-wide linkage analysis assessing parent-of-origin effects in the inheritance of birth weight. *Hum Genet* 2002;**110:**503–9.

doi:10.1093/ije/dym186
Advance Access publication 19 September 2007

Reply

From GEORGE DAVEY SMITH[1]* and SHAH EBRAHIM[2]

We thank Murielle Bochud and colleagues[1] for their letter regarding our two review articles on Mendelian randomization.[2,3] We agree with them that caution needs to be applied to the interpretation of the findings from Mendelian randomization studies. They raise two new issues that we would like to comment on. First, they suggest that transmission ratio distortion (TRD) could generate spurious interpretations. As we discussed in our first paper,[1] true Mendelian randomization would occur in studies relating to genetic variant transmission from both parents to their offspring and how this relates to offspring outcomes. TRD could indeed lead to erroneous conclusions in this setting. However, in practice, Mendelian randomization studies have relied on the generally well-supported notion that at a population level, genetic variants are unrelated to socioeconomic, behavioural and physiological confounding factors that would distort the interpretation of conventional epidemiological studies.[4–10] In this case, TRD would only lead to spurious conclusions if the distortion was influenced by a maternal characteristic (such as folate intake) that was socially patterned or demonstrated intergenerational behavioural or environmental continuity in such a way that this leads to genotype, at a population level, becoming associated with potential confounding factors. While claims have been made in this regard, we are not aware of robust evidence that such selective TRD occurs at a level that has led to associations between genetic variants and potential confounders at the population level.[10] However, we acknowledge that this is a possibility.

Second, Bochud *et al*. raised the issue of parent-of-origin effects. Again, in population level studies, this will only lead to erroneous conclusions if the process is

[1] Department of Social Medicine, University of Bristol, Bristol BS8 2PR, UK.

[2] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.

* Corresponding author.
E-mail: george.davey-smith@bristol.ac.uk