STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:2750–2755
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.3308

COMMENTARY

Is Mendelian randomization 'lost in translation?': Comments on 'Mendelian randomization equals instrumental variable analysis with genetic instruments' by Wehby *et al.*

Debbie A. Lawlor[1,*,†], Frank Windmeijer[1,2] and George Davey Smith[1]

[1]*MRC Centre for Causal Analysis in Translational Epidemiology, Department of Social Medicine, University of Bristol, Bristol BS8 2PR, U.K.*
[2]*Department of Economics, University of Bristol, Bristol BS8 2PR, U.K.*

Wehby *et al*. suggest that researchers use the term '*instrumental variable analysis with genetic instruments*' rather than '*Mendelian randomization*' for studies that use genetic variants that proxy for modifiable risk factors to investigate the causal effects of these risk factors [1]. They state that '*Using common language between disciplines applying IV analysis with genetic variants is essential to increasing collaborations and fostering the application of this method*' [1]. However, we would argue that truly translational research requires direct collaboration between individuals from different disciplines, clear definitions of discipline-specific terminology and the willingness to learn from each other. Any attempts to develop a common language must reflect this reality and may be less important than being able to translate between disciplines.

In the limited context of 'Mendelian randomization' developing the necessary multidisciplinary approach clearly goes well beyond the term used to describe these studies. What is required is capitalizing on the insights that all disciplines can bring and not arguing over which discipline should be the one with terminological hegemony. It is notable, for example, that throughout their commentary Wehby *et al*. refer to 'endogenous' variables, without any explanation of this term in language that would be understood by clinicians, epidemiologists, geneticists, basic scientists or indeed many biomedical statisticians for whom this term will be unknown. Similarly, they use the term 'direct' where most epidemiologists and clinicians would specify 'causal' (the latter also having a clearer understanding for lay readers of research).

Others have also suggested alternative terms for '*Mendelian randomization*' including '*Mendelian deconfounding*' [2] and '*Mendelian triangulation*' [3], so which name should be given prominence? We are not particularly committed to the name '*Mendelian randomization*' and have discussed the origin of this term in previous publications [4, 5]. However, it is now recognized

*Correspondence to: Debbie A. Lawlor, MRC Centre for Causal Analysis in Translational Epidemiology, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Rd, Bristol BS8 2PR, U.K.
†E-mail: d.a.lawlor@bristol.ac.uk

Copyright © 2008 John Wiley & Sons, Ltd.

within epidemiology, biostatistics, genetic epidemiology, molecular genetics and clinical practice and is listed and defined in Wikipedia (http://en.wikipedia.org/wiki/Mendelian_randomization). For example, it is included as a study type in the hierarchy of evidence (placed between prospective cohort studies and randomized controlled trials) in the most recent update of the American Academy of Pediatrics clinical guidelines.

Rather than trying to develop a common language we believe researchers should respect the expertise of disciplines different from their own and where appropriate work directly with them. Our concern here is emphasized by some studies that have used the term '*instrumental variable analysis with genetic instruments*,' or something very similar, but seem to us to have a poor understanding of the instruments that they are using (presumably from failure to work closely with relevant disciplines), and consequently may have made biased inference. For example, Brown *et al*. used family history of diabetes as a '*genetic instrumental variable*' (their term throughout the paper) to examine the causal association of diabetes with employment [6]. They assumed family history of diabetes to be unrelated to potential confounding factors of the association between diabetes and employment: '*Our IVs should be independent of factors*, *such as*, *for example*, *age at onset and comorbidities*' [6, p. 541]. For anyone with a clinical background this makes little sense since most cases of type II diabetes (the more common form) are asymptomatic, with up to 50 per cent remaining undiagnosed [7–9]. A family history of diabetes would result in detection bias, with family members and clinicians more likely to initiate screening for the disease than would be the case among individuals without diabetes. This would result in diagnosis at an earlier age in those with a family history than those without a family history. Furthermore, as risk of type II diabetes is related to social class [9], family history of diabetes will similarly be related to social class (and many factors that themselves associate with social class) and will certainly not serve as the kind of proxy marker suitable for use as an instrumental variable.

There are similar problems with the studies that Wehby *et al*. refer to: '*Some health economists have begun to apply IV methods with genetic variants as instruments to evaluate the effects of health risks and behaviours on health and socioeconomic outcomes*' [1]. It is now well established that many reported genetic associations are false positives in that they fail to replicate in larger independent studies [10]. We would only ever countenance using genetic variants in Mendelian randomization studies if they had clearly established robust associations with the risk factor of interest [11]. If this is not the case, then the basis of the Mendelian randomization study and the first assumption of any instrumental variables' analysis is violated [11]. We do not believe that any of the three studies cited by Wehby *et al*. [1] fulfil this basic criterion.

In the study by Ding *et al*., the authors acknowledge that in the literature there are inconsistencies in the associations between the genetic variants that they use and health or health-related behaviours [12, p. 33]. Indeed in their own study they test the association of four genetic variants with five health-related outcomes (one of which is split into two subgroups for some analyses) on a sample of fewer than 900 and fail to replicate most of the associations they initially claim as reported in the literature [12]. This is perhaps not surprising, given both the facts that the evidence of a robust association for any of these variants is lacking and the small sample size of their study. Furthermore, the authors fail to acknowledge the much more stringent requirements for very small p-values in genetic association studies than the conventional 5 per cent threshold [10]. Similarly, Norton and Han use a polymorphism in *DRD4* (the dopamine receptor gene) and one in *DAT1* (the dopamine transporter gene) as instrumental variables for obesity [13]. However, the association of these variants with body mass index or obesity has failed to replicate in large independent studies, with a recent review of the polygenetic basis of obesity

Copyright © 2008 John Wiley & Sons, Ltd.

confirming only two variants (*FTO* and *MC4R*) as having robust replicated associations [14]. We would not consider *DRD4*, *DAT1* or any other variant that has not been adequately replicated as a suitable instrument for obesity. Furthermore, Norton *et al.* claim a sex difference in the effect of these variants with body mass index, without recognition of the evidence demonstrating how spurious most claims of sex interactions with genetic effects are [15]. Finally, the cited example from Wehby's own team (publicly available only in abstract form and therefore difficult to fully evaluate) claims effects of five individual single nucleotide polymorphisms (SNPs) on smoking to have relative risks of 2.0–3.0 [16]. The magnitudes of these associations are difficult to believe and we would be cautious of accepting the author's IV results without further replication of these associations in large independent samples. For example, a previous publication for one of the variants found to be associated with smoking by Wehby *et al.* (*DBH*) found no association with smoking behaviour or cotinine levels (a biomarker that accurately measures smoking without reporting bias) [17].

One further problem with the suggested name change to '*instrumental variable analysis with genetic instruments*' is that several published examples fulfil our definition of Mendelian randomization, in that they use knowledge of a robust association between a genetic variant and a risk factor of interest to provide evidence regarding the causal association of that risk factor with an outcome, but they do not use a formal instrumental variables' analysis. In these studies information on the association between the genetic variant and the risk factor of interest is obtained from different study populations to those that provide information on the association between the genetic variant and the outcome of interest.

For example, in a study examining the association between c-reactive protein (CRP) and coronary heart disease (CHD), the authors calculated the weighted mean difference in CRP between individuals with variants of the $+1444C>\mathrm{T}$ polymorphism in the *CRP* gene among European individuals from six studies [18]. They then used those results together with data from previously published observational studies to compute an expected odds ratio for CHD among individuals homozygous for the T allele (expected if the observational study results were unbiased). The authors then performed four new genetic association studies (in European populations) to obtain an odds ratio for the association between the $+1444C>\mathrm{T}$ polymorphism and CHD. The expected odds ratio for CHD by genotype was 1.20 (95 per cent CI 1.07–1.38), whereas the observed odds ratio was 1.01 (95 per cent CI 0.74–1.38), leading the authors to conclude that the observational studies appeared to exaggerate the causal effect of CRP [18].

In a recent Mendelian randomization study, a meta-analysis of all published studies of the association between *aldehyde dehydrogenase 2* (*ALDH2*), a genetic variant that encodes a major enzyme involved in alcohol metabolism, which is robustly associated with levels of alcohol consumption, and blood pressure was undertaken [19]. A marked difference in mean blood pressure (7.44 mmHg (95 per cent CI 5.39, 9.49), $p=1.1\times10^{-2}$) between wild-type homozygotes ($*1*1$), who would not have any adverse consequences from consuming alcohol, and those homozygous for the null allele ($*2*2$), who would have an adverse effect to alcohol consumption and who consume much lower levels of alcohol, was found. These findings provide strong evidence that alcohol intake has a marked causal effect on blood pressure [19]. No formal instrumental variables' analysis was possible since individual participant data were not available and not all studies included in the meta-analysis had information on alcohol consumption.

Although we note in our earlier *Statistics in Medicine* publication the potential limitations of using information on gene-risk factor association and gene-outcome associations from different studies [11], we nonetheless believe that such studies (with clear biological understanding and

Copyright © 2008 John Wiley & Sons, Ltd.

explicit assumptions) can, and do, make important contributions to the literature. Despite the lack of formal instrumental variables' analysis these fulfil our criteria for Mendelian randomization studies.

Wehby *et al*. go on to discuss the importance of testing (as far as one can) for underlying instrumental variables' assumptions and having clear rationales for undertaking an instrumental variables' analysis and for the choice of instruments used. We agree completely with these principles [11]. However, we disagree with the authors' assertion that in our earlier publication [11] we failed to describe any tests for assessing whether linkage disequilibrium and pleiotropy resulted in possible violation of the instrumental variables' assumption. We do indeed discuss how two or more genetic variants with known different biological pathways to the risk factor of interest could be used to test this possible violation [11]. It is true, however, that when there is only one known genetic variant the possible violation cannot be tested, but this fact underlies the use of instrumental variables, analysis throughout econometrics and other disciplines. As Cox and Wermuth point out '*the special independence assumptions made in formulating these equations cannot be empirically tested*' [20, p. 297] and therefore argue that instruments should be selected '*only on subject-matter grounds*' [20, p. 297] (i.e. best scientific evidence). It is therefore essential to use biological (or other relevant) knowledge in instrument selection for studies in order to make correct causal inference in instrumental variables' analysis.

We acknowledge that the intricacies of biological pathways make it almost always possible to invoke a biological justification for an epidemiological (including a genetic epidemiological) association. For these reasons we believe that the selection of genetic variants for use in Mendelian randomization studies must be guided by the strength of evidence regarding any robust association of the genetic variant with the risk factor of interest. Our criteria here are driven by demonstrating consistent associations in a number of very large independent studies and applying much more stringent p-values than conventionally used in medical statistics [10].

We are also unsure why Wehby *et al*. suggest that '*the endogenous selection of biologic markers or exposures such as plasma folate levels, C-reactive proteins or others is less intuitive*.' These issues have been discussed fully elsewhere [21–24]. In brief, numerous claims are made, almost on a daily basis, for the causal association of such exposures with disease outcomes. These claims are most commonly made on the basis of associations found in observational epidemiological studies, sometimes prospective studies, but commonly prevalence (case control or cross-sectional) studies (for example, in a recent review we identified 81 prevalent studies of the association between CRP and various types of cancer, most claiming a positive association, but only nine prospective studies [25]). Such associations are likely to be related to confounding and/or reverse causality and yet they are used as a rationale to begin developing and evaluating drugs that affect the biologic exposures. Genetic variants that are robustly associated with these biologic exposures are unlikely to be associated with characteristics that commonly confound associations of the exposure with disease outcome [26], and the association of the genetic variant with the outcome cannot be explained by reverse causality [11]. Thus, Mendelian randomization studies of biological exposures are a useful tool for providing valid causal inference. They also have additional advantages in that they provide an estimate of the causal effect of lifelong mean differences in the biological exposure. For example, the associations of SNPs, which are known to be robustly associated with variation in low-density lipoprotein cholesterol (i.e. associated in two or more very large independent studies with appropriately very small p-values), with CHD outcomes are greater than would be expected from the known effect of random allocation to statin medication (which acts via a reduction in low-density lipoprotein cholesterol) [27]. The most likely explanation for this greater effect is that

Copyright © 2008 John Wiley & Sons, Ltd.

the variants are related to lifetime differences in low-density lipoprotein cholesterol, whereas statin use in mid-life by definition only affects levels from this time onwards [28, 29].

Wehby *et al*.'s comment on the importance of knowing just how genetic variants affect risk factors, in particular health behaviours, is very important and is discussed in our original commentary [11]. Here a useful analogy for clinicians, epidemiologists and health service researchers is with the randomization to advice to change a health behaviour. Results from an intention to treat analysis of a randomized controlled trial showing a change in a disease outcome by randomization to dietary advice, for example, are correctly interpreted as showing the effect of the dietary advice (and not necessarily the diet *per se*).

In conclusion, we agree with Wehby *et al*. [1] that all branches of science can be enhanced by greater collaboration across relevant disciplines. However, we feel that this requires direct working with each other, and by this collaborative work gaining knowledge of differences (and similarities) in language, philosophy, understanding and methods, rather than trying to develop a single common language.

REFERENCES

1. Wehby G, Ohsfeldt R, Murray J. Mendelian randomization equals instrumental variable analysis with genetic instruments. *Statistics in Medicine* 2008; DOI: 10.1002/sim3255.
2. Tobin MD, Minelli C, Burton PR, Thompson JR. Commentary: development of Mendelian randomization: from hypothesis test to 'Mendelian deconfounding'. *International Journal of Epidemiology* 2004; **33**:26–29.
3. Bautista LE, Smeeth L, Hingorani AD, Casas JP. Estimation of bias in nongenetic observational studies using 'Mendelian triangulation'. *Annals of Epidemiology* 2006; **16**:675–680.
4. Davey Smith G. Capitalizing on Mendelian randomization to assess the effects of treatments. *Journal of the Royal Society of Medicine* 2007; **100**:432–435.
5. Davey Smith G. Capitalising on Mendelian randomisation to assess the effects of treatment. *James Lind Library* 2006; **100**:432–435.
6. Brown III HS, Pagan JA, Bastida E. The impact of diabetes on employment: genetic IVs in a bivariate probit. *Health Economics* 2005; **14**:537–544.
7. Williams R, Wild S. Diabetes prevalence data for the United Kingdom—what do we have and what do we need? *Diabetic Medicine* 2003; **20**:505–506.
8. Thomas MC, Walker MK, Emberson JR, Thomson AG, Lawlor DA, Ebrahim S, Whincup PH. Prevalence of undiagnosed type 2 diabetes and impaired fasting glucose in an older British population. *Diabetic Medicine* 2005; **22**:789–793.
9. Lawlor DA, Patel R, Fraser A, Davey Smith G, Ebrahim S. The association of life course socio-economic position with diagnosis, treatment, control and survival of women with diabetes: findings from the British Women's Heart and Health Study. *Diabetic Medicine* 2007; **24**:892–900.
10. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes: can we avoid being swamped by spurious findings? *Lancet* 2003; **361**:865–872.
11. Lawlor DA, Harbord RM, Sterne JAC, Timpson NJ, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* 2007; DOI: 10.1002/sim.3034.
12. Ding W, Lehrer SF, Rosenquist JN, Audrian-McGovern J. The impact of poor health on education: new evidence using genetic markers. *NBER Working Paper Series 2006*; *Working Paper 12304*, 2006. Available from: www.nber.org/papers/w12304.
13. Norton EC, Han E. Genetic information, obesity, and labor market outcomes. *HEDG Working Paper Series 2007*; *Working Paper 07/15*, 2007. Available from: www.york.ac.uk/res/herc/hedgwp.
14. Hinney A, Hebebrand J. Polygenic obesity in humans. *Obesity Facts—The European Journal of Obesity* 2008; **1**:35–42.
15. Patsopoulos NA, Tatsioni A, Ioannidis JP. Claims of sex differences: an empirical assessment in genetic associations. *Journal of the American Medical Association* 2007; **298**:880–893.
16. Wehby GL, Marcinow A, Quin X, Mansilla A, Murray JC. A genetic instrumental variables analysis of the effects of maternal smoking on oral cleft risks. *Fifty-seventh Annual Meeting of the American Society of Human Genetics*, San Diego, CA, U.S.A., 2007 (Abstract).

Copyright © 2008 John Wiley & Sons, Ltd.

17. Huang S, Cook DG, Hinks LJ, Chen XH, Ye S, Gilg JA, Jarvis MJ, Whincup PH, Day IN. CYP2A6, MAOA, DBH, DRD4, and 5HT2A genotypes, smoking behaviour and cotinine levels in 1518 UK adolescents. *Pharmacogenetics and Genomics* 2005; **15**:839–850.
18. Casas JP, Shah T, Cooper J, Hawe E, McMahon AD, Gaffney D, Packard CJ, O'Rielly DS, Juhan-Vague I, Judkin JS, Tremoli E, Margaglione M, Di Minno G, Hamsten A, Kooistra T, Stephens JW, Hurel SJ, Livinbstone S, Colhoun HM, Miller GJ, Bautista LE, Meade T, Sattar N, Humphries SE, Hingorani AD. Insights into the nature of the CRP-coronary event association using Mendelian randomization. *International Journal of Epidemiology* 2006; **35**:922–934.
19. Chen L, Davey Smith G, Harbord RM, Lewis SJ. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLoS Medicine* 2008; **5**:e52. DOI: 10.1371/journal.pmed.0050052.
20. Cox DR, Wermuth N. Causality: a statistical view. *International Statistical Review* 2004; **72**:285–305.
21. Davey Smith G, Ebrahim S. Mendelian randomisation: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**:1–22.
22. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Rumley A, Lowe GDO, Day INM, Ebrahim S. Association of C-reactive protein with blood pressure and hypertension: lifecourse confounding and Mendelian randomisation tests of causality. *Arteriosclerosis Thrombosis and Vascular Biology* 2005; **25**:1051–1056.
23. Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day IN, Palmer LJ, Hattersley AT, Ebrahim S, Lowe GD, Rumley A, Davey Smith G. C-reactive protein and its role in metabolic syndrome: Mendelian randomisation study. *Lancet* 2005; **366**:1954–1959.
24. Lewis S, Lawlor DA, Davey Smith G, Araya R, Timpson N, Day IMN, Ebrahim S. The thermolabile variant of MTHFR is associated with depression in the British Women's Heart and Health Study and a meta-analysis. *Mol Psychiatry* 2006; DOI: 10.1038/sj.mp.4001790.
25. Heikkilä K, Ebrahim S, Lawlor DA. Systematic review of the association between circulating interleukin-6 (IL-6) and cancer. *European Journal of Cancer* 2008; DOI: 10.1016/j.ejca.2008.02.047.
26. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes a fundamental distinction between conventional and genetic epidemiology. *PLoS Medicine* 2008; **4**:e352. DOI: 10.1371/journal.pmed.0040352.
27. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. *New England Journal of Medicine* 2007; **357**:443–453.
28. Krauss RM. What can the genome tell us about LDL cholesterol? *Lancet* 2008; **371**:450–452.
29. Brown MS, Goldstein JL. Biomedicine. Lowering LDL—not only how low, but how long? *Science* 2006; **311**:1721–1723.

Please see the related Letter to the Editor and Authors' Reply on pages 2974–2978.

Copyright © 2008 John Wiley & Sons, Ltd.