outcome than a direct intervention because of the specific pathway affected by the variant, such as the effect of kringle IV type 2 size polymorphisms on lipoprotein (a), and the subsequent association with myocardial infarction(5).

While statistical guidance on assessment of the validity of IVs in Mendelian randomization is welcome (6), there is a danger of overreliance on empirical testing at the expense of biologic knowledge (7). The statements provided by Glymour et al., while providing useful guidance, should not be seen as absolute indicators of the invalidity of an IV and should supplement rather than replace sound scientific judgment (8).

ACKNOWLEDGMENTS

The author is supported by the British Heart Foundation (program grant RG/08/014).

Conflict of interest: none declared.

REFERENCES

1. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol*. 2012;175(4):332–339.
2. Kivimäki M, Jokela M, Hamer M, et al. Examining overweight and obesity as risk factors for common mental disorders using fat mass and obesity-associated (*FTO*) genotype-instrumented analysis: the Whitehall II Study, 1985–2004. *Am J Epidemiol*. 2011;173(4):421–429.
3. Davey Smith G. Randomised by (your) god: robust inference from an observational study design. *J Epidemiol Community Health*. 2006;60(5):382–388.
4. Smith G, Timpson N, Ebrahim S. Strengthening causal inference in cardiovascular epidemiology through Mendelian randomization. *Ann Med*. 2008;40(7):524–541.
5. Kamstrup P, Tybjaerg-Hansen A, Steffensen R, et al. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA*. 2009;301(22):2331–2339.
6. Wehby G, Ohsfeldt R, Murray J. "Mendelian randomization" equals instrumental variable analysis with genetic instruments. *Stat Med*. 2008;27(15):2745–2749.
7. Lawlor DA, Windmeijer F, Davey Smith G. Is Mendelian randomization 'lost in translation?': comments on 'Mendelian randomization equals instrumental variable analysis with genetic instruments' by Wehby et al. *Stat Med*. 2008;27(15):2750–2755.
8. von Hinke Kessler Scholder S, Davey Smith G, Lawlor DA, et al. Mendelian randomization: the use of genes in instrumental variable analyses. *Health Econ*. 2011;20(8):893–896.

Stephen Burgess (e-mail: sb452@medschl.cam.ac.uk)
Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge CB1 8RN, United Kingdom

DOI: 10.1093/aje/kws249; Advance Access publication: July 31, 2012

In their recent discussion evaluating the instrumental variable assumptions in Mendelian randomization studies, Glymour et al. (1) discussed Bonet's instrumental inequalities (2), which can be applied when the instrumental variable, exposure, and outcome are all binary. In their Web Appendix 3, they presented an example from Kivimäki et al. (3) in which the effect of being obese (body mass index (weight (kg)/height (m)^2) $\geq$30) on the risk of common mental disorders was estimated using genotypes of the rs1421085 polymorphism in the fat mass and obesity-associated gene (*FTO*) as an instrumental variable. In the example, Glymour et al. found that the instrumental inequalities were satisfied, and they calculated bounds for the probability $P(Y(X=0)$ of the exposure-free potential outcome of 3%–17% (where $Y(X=x)$ denotes the potential outcome (Y), given that an exposure (X) is set to some value x).

Further to this, we point out that Balke and Pearl (4) derived nonparametric bounds for the average causal effect (ACE), where the ACE is the difference between $P(Y(X=1))$ and $P(Y(X=0))$. We have implemented the inequality check and the calculation of such bounds, as well as several extensions to it, in a Stata command (StataCorp LP, College Station, Texas) called bpbounds (5). The extensions allow for an instrument with 3 levels, for data on exposure and outcome coming from separate studies, and we illustrate the calculations also for case-control data. For the Kivimäki et al. data example used by Glymour et al., we find that the bounds for the ACE are (−16%, 75%) (Table 1). Note that such bounds are interpreted differently than confidence intervals. They tell us that there exists some distribution involving the unobserved confounders (between the exposure and outcome) that yields a true ACE as small as −16%, while another existing distribution involving the unmeasured confounders has a true ACE as large as 75%, with both distributions satisfying the instrumental variable assumptions and having the same observed marginal frequencies for the exposure, outcome, and instrument. Since the confounder is unobserved, it is impossible to decide where the ACE lies in the interval from the observable data without making further parametric assumptions such as those made in linear models. In this particular example, calculation of the bounds shows that the data alone are not especially informative for the causal effect, as the bounds are wide and include the case of no causal effect. However, it is important to realize that any additional analysis (based, for example, on a linear model) only produces a precise point estimate because of such additional assumptions. Because such parametric assumptions are often difficult to justify, we recommend always supplementing such analyses with the above nonparametric bounds for the causal effect. Note that the instrumental inequalities assuming a monotonic relation between the instrument and exposure are not satisfied in the Kivimäki et al. data example.

Importantly, checking the inequalities is not a statistical test in the usual sense; for example, it is possible for the instrumental variable assumptions to not be met by a specific data-generating process but the instrumental inequalities to be satisfied (for an example of this, see section 8.3 of the article by Palmer et al. (5)). In addition, solely calculating

the bounds does not take into account sampling variability. To address this, Ramsahai and Lauritzen (6) proposed a relevant hypothesis test for the bounds. In related work, Richardson et al. (7) also proposed a Bayesian approach to estimating bounds for the ACE and other causal parameters.

In conclusion, when the exposure and outcome in a Mendelian randomization analysis are binary variables and the instrument is a categorical variable, gross violations of the instrumental variable assumptions, including the exclusion restriction, can sometimes be detected by checking certain inequality restrictions on the observed relative frequencies. Further empirical evidence for violations of the instrumental variable assumptions can sometimes be obtained using multiple instruments and overidentification tests (8). However, we caution researchers that it is generally not possible to establish the *validity* of the instrumental variable assumptions, particularly the exclusion restriction assumption, on the basis of data and statistical tests alone. In general, the exclusion restriction should always be justified from subject matter background knowledge—in this example, the biochemical and behavioral mechanisms underlying the *FTO* gene (9).

ACKNOWLEDGMENTS

Conflict of interest: none declared.

This work was supported by United Kingdom Medical Research Council grants G0601625 and G0600705 and Leverhulme Trust grants RF/9/RFG/2009/0062 and RF-2011-320.

REFERENCES

1. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol*. 2012;175(4):332–339.
2. Bonet B. Instrumentality tests revisited. In: Breese JS, Koller D, eds. *UAI '01: Proceedings of the Seventeenth Annual Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufman, Inc; 2001:48–55.
3. Kivimäki M, Jokela M, Hamer M, et al. Examining overweight and obesity as risk factors for common mental disorders using fat mass and obesity-associated (*FTO*) genotype-instrumented analysis: the Whitehall II Study, 1985–2004. *Am J Epidemiol*. 2011;173(4):421–429.
4. Balke A, Pearl J. Bounds on treatment effects from studies with imperfect compliance. *J Am Stat Assoc*. 1997;92(439):1171–1176.
5. Palmer TM, Ramsahai RR, Didelez V, et al. Nonparametric bounds for the causal effect in a binary instrumental-variable model. *Stata J*. 2011;11(3):345–367.
6. Ramsahai RR, Lauritzen SL. Likelihood analysis of the binary instrumental variable model. *Biometrika*. 2011;98(4):987–994.
7. Richardson TS, Evans RJ, Robins JM. Transparent parameterizations of models for potential outcomes. In: Bernardo JM, Bayarri MJ, Berger JO, eds. *Bayesian Analysis 9*. New York, NY: Oxford University Press; 2010:569–610.
8. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res*. 2012;21(3):223–242.
9. Timpson NJ, Emmett PM, Frayling TM, et al. The fat mass- and obesity-associated locus and dietary intake in children. *Am J Clin Nutr*. 2008;88(4):971–978.

Tom M. Palmer[1,2], Roland R. Ramsahai[3], Debbie A. Lawlor[1,2], Nuala A. Sheehan[4,5], and Vanessa Didelez[6] (e-mail: tom.palmer@bristol.ac.uk) (e-mail: tom.palmer@bristol.ac.uk)

[1] *MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol BS8 2BN, United Kingdom*
[2] *School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, United Kingdom*
[3] *Statistical Laboratory, University of Cambridge, Cambridge CB3 0WB, United Kingdom*
[4] *Department of Health Sciences, University of Leicester, Leicester LE1 7RH, United Kingdom*
[5] *Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom*
[6] *School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom*

DOI: 10.1093/aje/kws250; Advance Access publication: July 31, 2012

THE AUTHORS REPLY

We appreciate the comments of Palmer et al. (1) and Burgess (2), as well as Palmer et al.'s provision of their useful Stata command, bpbounds (StataCorp LP, College Station, Texas). As Palmer et al. note (1), the assumptions required for a valid instrumental variable (IV) cannot be established from any data or statistical tests, but these assumptions can sometimes be falsified. One approach to falsification, applying the instrumental inequality tests, is applicable with categorical IVs and phenotypes (3, 4). As we demonstrated in Web Appendix 3 of our article (5), these tests are straightforward to implement in Excel (Microsoft Corporation, Seattle, Washington) when the phenotype is dichotomous and the instrument is either dichotomous or trichotomous. Trichotomous instruments are particularly common in Mendelian randomization studies. In our example, we used *FTO* allele count as a trichotomous instrument, classified as homozygous for the common allele, heterozygous, or homozygous for the rare allele.

A valid IV places inequality constraints on the observed data distribution. An instrumental inequality test assesses whether these constraints hold in the data. The ability (i.e., power) of the test to detect an invalid instrument increases with the number of constraints being tested; hence, it is optimal to test all of the inequality constraints. Pearl (6) first derived constraints implied by a valid IV; Bonet (4) subsequently recognized that when either the IV or the endogenous variable has more than 2 possible values, additional inequality constraints are implied by the IV assumptions. Bonet demonstrated how to derive and thus test all inequality constraints; Bonet's tests are implemented in Web Appendix 3 of our original article (5) and in the bpbounds command. In Web Figure 1 (available at http://aje.oxfordjournals.org/),