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

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/),

we provide an example of a hypothetical distribution for a proposed trichotomous IV ($Z$) and dichotomous exposure ($X$) and outcome ($Y$) which violates Bonet’s tests but not Pearl’s.

Although violation of the instrumental inequality constraints implies that the proposed IV is not a valid instrument for the exposure and outcome examined, lack of violation does not imply that the IV is valid. Even extreme violations of the IV assumptions, large enough to produce severely biased IV estimates, may not violate the inequality constraints. An important caution in IV analyses is that small violations of the assumptions can introduce biases much larger than similar violations would typically induce in conventional analyses. Further, if the inequality constraints are violated, one explanation is that a continuous causal phenotype was inappropriately dichotomized to define the exposure. For example, consider the situation in which we use “obesity” (body mass index (BMI; weight (kg)/height (m)$^2$) $\geq 30$) as the exposure, but in fact the causally relevant exposure is BMI, which has a continuous dose-response relation with the outcome. A variable which is a valid IV for the effect of BMI on the outcome may nonetheless violate the instrumental inequality constraints when obesity is used as the phenotype. This is consistent with the fact that the IV, though valid for estimating the effects of BMI, is not valid for estimating the effects of obesity.

We agree with Palmer et al. that one should routinely provide nonparametric bounds for the average causal effect (and confidence intervals for these bounds). Their Stata macro provides such bounds for a dichotomous $X$ and a dichotomous or trichotomous $Z$. Finally, we are in agreement with the remainder of the Palmer et al. letter, which reemphasizes several points we made in our article.

In our article (5), we showed that if researchers are willing to assume a linear model and to make an assumption about the direction of confounding between an exposure ($X$) and an outcome ($Y$), there are 4 equivalent tests for a valid IV. These were described in statements 1–4 in our article, which should properly refer to population estimates, such as regression coefficients, not sample estimates. Thus, we agree with Dr. Burgess’ point (2) that, because of sampling variability, a valid IV may fail the tests and an invalid IV may pass the tests. As is common, we assessed the plausibility of sampling variability as an explanation for our findings by reporting a $P$ value. In particular, the $P$ value for the test that the correlation between $Z$ and $Y$ given $X$ was nonpositive in the Kivimäki et al. example (7) suggested that sampling variability was an unlikely explanation.

ACKNOWLEDGMENTS

Financial support was provided by the National Institute of Mental Health (grant MH092707-01) and the National Institute of Environmental Health Sciences (grant ES019712-01).

Conflict of interest: none declared.

REFERENCES


M. Maria Glymour$^1$, Eric J. Tchetgen Tchetgen$^2$, and James M. Robins$^2$ (e-mail: mglymour@hsph.harvard.edu)

$^1$Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA 02115
$^2$Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115

DOI: 10.1093/aje/kws251; Advance Access publication: July 31, 2012