

to be robust) rely on the results of these analyses in isolation, and indeed many of the other analyses reported in our paper do not suffer from potential bias due to weak instruments.

Hartwig and Davies did, however, suggest that we could have used estimates from an external source to obtain less biased results in our MR-Egger analyses. Whereas we agree that this would be good practice in most situations, we do not feel that it would have been appropriate in our study, for two reasons. First, the focus of our article was not on a possible causal relationship between body mass index and BMD (which is well-known and widely accepted), but rather on a possible causal relationship between adiposity [as operationalized as fat mass calculated from total body dual-energy X-ray absorptiometry (DXA)] and BMD. There are no publicly available genome-wide association studies of total body fat mass as measured by total body DXA, and therefore no external estimates that we could have applied in our analyses (i.e. as far as we are aware, we are currently the largest such study). We could have used external estimates for analyses involving body mass index, but this

would have been of limited utility since body mass index is a far from perfect measure of adiposity. Second, our study involved 9-year-old children from the Avon Longitudinal Study of Parents and Children. It is unclear the extent to which effect sizes of adiposity-associated variants in adults reflect effect sizes of adiposity-associated variants in children (as Hartwig and Davies recognize), and we therefore feel it would have been inappropriate to use adult-derived external estimates in our study of children.

## References

1. Hartwig F, Davies N. Why internal weights should be avoided (not only) in MR Egger regression. *Int J Epidemiol* 2016;**45**:1676–78.
2. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44**:512–25.
3. Kemp JP, Sayers A, Davey Smith G, Tobias JH, Evans DM. Using Mendelian randomization to investigate a possible causal relationship between adiposity and increased bone mineral density at different skeletal sites in children. *Int J Epidemiol* 2016;**45**: 1560–72.

## Response to Hartwig and Davies

From Jack Bowden,<sup>1,2\*</sup> Stephen Burgess<sup>3</sup> and George Davey Smith<sup>1</sup>

<sup>1</sup>MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK, <sup>2</sup>MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK and <sup>3</sup>Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

\*Corresponding author. MRC Integrative Epidemiology Unit, University of Bristol, School of Social and Community Medicine, Oakfield Grove, Bristol BS8 2BN, UK. E-mail: jack.bowden@bristol.ac.uk

*International Journal of Epidemiology*, 2016, 1679–1680

doi: 10.1093/ije/dyw252

Advance Access Publication Date: 20 September 2016



Thank you for the opportunity to respond to the thoughtful and timely letter by Hartwig and Davies.<sup>1</sup> They do indeed raise a very important practical issue with the implementation of MR-Egger regression in the single-sample setting, or with the use of so called ‘internal’ weights, namely weak instrument bias. They are right to point out the unsatisfactory nature of our analysis of the height data in our original publication,<sup>2</sup> and that our use of weak instruments had the likely effect of biasing the MR-Egger estimate towards that of the observational association. Their re-analysis of these data with external weights appears to provide a much more satisfactory answer and, when such weights are available, it is both a simple and an attractive way to circumvent the problem.

Although previous simulation studies have highlighted this fact, further research is needed to completely understand the issue of weak instrument bias for MR-Egger in

the single-sample context. What is clear however, is that the standard notion of instrument strength, as quantified by the F statistic, cannot naively be applied to estimate the magnitude of this bias; new (or at least newly borrowed) theory is required. Before covering initial progress in this vein, we now briefly discuss related (and more mature) work in the two-sample context.

## Recent work on weak instrument bias in the two-sample context

A strength of MR-Egger regression, along with the weighted median<sup>3</sup> and inverse-variance weighted (IVW)<sup>4</sup> methods, is that they do not require gene, exposure and outcome data on a single sample of subjects at the individual level. MR-Egger regression can be implemented with only summary data estimates of the single nucleotide polymorphism (SNP)-exposure and SNP-outcome associations, making it most

natural to use within the two-sample summary data context. This is becoming an increasingly popular way of performing Mendelian randomization investigations and, in recent work to appear in this journal,<sup>5</sup> we extensively explore the issue of weak instrument bias for MR-Egger in this simpler setting. We show that weak instruments (with small F statistics) have the effect of inducing regression dilution bias into the MR-Egger estimate, thus shrinking its value towards the null (although such bias does not inflate Type 1 error rates when testing for a causal effect). This same phenomenon has also been shown for the IVW estimate.<sup>6,7</sup> However, whereas the magnitude of the attenuation in the IVW estimate can be gauged from the instrument's F statistics (e.g. a mean F of 20 would indicate an approximate dilution of  $1/20 = 5\%$ ), this will almost certainly underestimate the dilution for MR-Egger. A new statistic - which is a simple modification of Higgins'  $I^2$  - can be used to quantify the dilution (we call it  $I^2_{GX}$ ). It lies between 0 and 100%, and an  $I^2_{GX}$  of 95% would indicate an expected dilution of 5%. We also describe how the established method of simulation extrapolation (SIMEX)<sup>8</sup> can be used to calculate a bias-adjusted MR-Egger estimate to mitigate the effect of this dilution.

### Preliminary results and further work in the single-sample context.

The first author of this letter is currently conducting a theoretical investigation of the weak instrument bias of MR-Egger regression in the single-sample setting, with particular focus on comparing its performance to that of the IVW estimate. Initial results highlight the following: (i) when all genetic instruments are valid, both MR-Egger and IVW estimates are consistent, but the MR-Egger estimate will always be more strongly affected by weak instrument bias (towards the observational association) than the IVW estimate; and (ii) when some or all of the genetic instruments are invalid due to directional pleiotropy, but the sample covariance of the pleiotropy and instrument strength terms is zero (e.g. the InSIDE assumption in Bowden *et al.*<sup>2</sup> holds), then MR-Egger is a consistent estimate of the causal effect whereas the IVW estimate is asymptotically biased. However, the finite-sample bias of the IVW estimate may still be less than that of MR-Egger if the genetic variants are sufficiently weak.

Research is under way to investigate whether a suitable statistic, perhaps analogous to  $I^2_{GX}$ , can be derived to quantify the likely magnitude of dilution of MR-Egger regression in the single-sample setting. An obvious follow-on question will be whether the SIMEX approach outlined in this journal<sup>5</sup> can also be transferred to this setting to yield a bias-adjusted estimate. This is not trivial because such an implementation is likely to require at least some a priori

knowledge of the strength of correlation between the standard errors of the SNP-exposure and SNP-outcome association estimates due to shared confounding (in the two-sample context we assume *a priori* this correlation is zero). Theoretical work on this topic already exists<sup>9</sup> and will no doubt serve as a useful starting point.

Even in cases which appear superficially to be two-sample settings, sample overlap may mean that the bias due to weak instruments is more similar to the single-sample setting. For example, around 71% of participants are shared between the Genetic Investigation of Anthropometric Traits (GIANT) consortium<sup>10</sup> and the Global Lipids Genetics Consortium (GLGC).<sup>11</sup> For the IVW method, the direction of weak instrument bias varies linearly as the proportion of sample overlap increases from the two-sample setting (where bias is in the direction of the null) to the single-sample setting (where bias is in the direction of the observational association).<sup>12</sup> Further work is needed to see if a similar pattern holds for the MR-Egger method.

### References

- Hartwig F, Davies N. Why internal weights should be avoided (not only) in MR Egger regression. *Int J Epidemiol* 2016;45:1676–78.
- Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512–25.
- Bowden J, Davey Smith G, Haycock P, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;40:304–14.
- Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37:658–65.
- Bowden J, Del Greco FM, Minelli C *et al.* Assessing the validity of MR-Egger regression for two-sample Mendelian randomization: the role of the  $I^2$  statistic. *Int J Epidemiol* 2016; doi: 10.1093/ije/dyw220.
- Angrist JD, Krueger AB. Split-sample instrumental variables estimates of the return to schooling. *J Bus Econ Stat* 1995;13: 225–35.
- Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol* 2013;178:1177–84.
- Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. *J Am Stat Assoc* 1994;89:1314–28.
- Ronning G, Rosseman M. SIMEX estimation in case of correlated measurement errors. *Adv Stat Anal* 2008;92:391–404.
- Locke AE, Kahali B, Berndt SI *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;518:197–206.
- Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45:1274–83.
- Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016; doi: 10.1002/gepi.21998.