Commentary. Fixed effects and risks of miscommunication: a comment on Jullien, Sinclair, and Garner

Owen Ozier

Development Research Group, Human Development and Public Services Team, The World Bank

Corresponding author. MSN MC3-311, 1818 H Street NW, Washington, DC 20433, USA. E-mail: oozier@worldbank.org

Disclaimer: The findings, interpretations, and conclusions expressed in this paper are entirely those of the author. They do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent.

Note: this article is published in the International Journal of Epidemiology; the final version may be accessed online at one of these locations:

https://academic.oup.com/ije/article/doi/10.1093/ije/dyw349/2970178/Commentary-Fixed-effects-and-risks-of

https://doi.org/10.1093/ije/dyw349

Jullien *et al.* (henceforth, JSG) [1] frame their manuscript as an appraisal of all long-term follow-up studies of mass deworming. This is a welcome and important undertaking. It becomes another exercise, however, when the search criteria reveal that the only such studies are those written by social scientists rather than public health or medical researchers. The result is an examination of the work of social scientists by public health researchers. This examination includes my own work (Ozier 2016), which JSG discuss in some detail.[2] Because it is under revision with a peer-reviewed journal, my draft manuscript has benefited from our correspondence. Nonetheless, some details of Ozier 2016 are misrepresented in the JSG discussion, exemplifying the pitfalls that may complicate such interdisciplinary exchange. The main critiques JSG present of Ozier 2016 are inaccurate, as I discuss below. Thus, JSG's conclusion that Ozier 2016 is "at risk of substantial methodological bias" is incorrect.

Ozier 2016 finds positive effects of deworming early in life on cognitive measures, but no effects of deworming on stature. Of Ozier 2016, JSG write, *"positive effects are taken from analyses across the whole sample, which include non-randomized data."* This is incorrect. A substantive misunderstanding here concerns the use of "fixed effects" in regression specifications identified from randomized variation of the introduction of deworming. Fixed effects in stepped wedge designs, used in the manner I describe below, are well known in the medical field; the explanation I provide here is not unique to economics.[3] The Ozier 2016 design works as follows. In an earlier study by Miguel and Kremer, the year in which deworming began in a community was randomized: 1998, 1999, or 2001.[4] Ozier 2016 examines children in those communities in relation to an indicator for deworming spillovers (deworming of older children) by age 1, a decision motivated by literature on the importance of health in sensitive periods of early childhood.[5] However, being exposed to deworming spillovers before age 1 is correlated with simply *being younger*, since all the very young birth cohorts in the study were born in places where deworming was already taking place. Within the 1998 and 1999 birth cohorts, however, whether children were exposed to deworming spillovers by age 1 is randomized. Fixed effects allow Ozier 2016 to pool this estimation across cohorts: conditional on the inclusion of indicator variables for each birth cohort (in other words, "fixed effects"), deworming by age 1 is randomly determined because the start date of deworming was randomized across communities. As Ozier 2016 describes the analysis, in addition to an indicator for treatment by age 1, *"fixed effects for the interaction of age, sex, and data collection year are also included."* The original analysis thus provides unbiased estimates derived from variation in the introduction of deworming that is random conditional on cohort. JSG write, *"Analyses across the whole sample (seven cohorts) are thus secondary observational analyses, with unknown* *secular changes potentially confounding the findings."* This is false. Cohort fixed effects remove any secular changes.

Additional missteps follow from that one substantive misunderstanding. JSG write, *"additional tables were produced confirming these effects were still apparent in analyses limited to the quasi-randomized cohorts from 1998 and 1999."* This is almost mechanically true since those two cohorts contain the relevant variation in treatment status. JSG also write, *"the revised analysis is substantially underpowered to reliably detect these effects…"* This is incorrect. The analysis to which they refer has power 0.84 to detect the coefficient estimated in that specification. Less substantively, but as a matter of record, it is also false that *"authors confirmed that only the analyses including all seven annual cohorts were adequately powered."* Confusion over our correspondence arises because JSG *earlier* wrote to ask whether separate group-cohort comparisons (Ozier 2016 Figure 1, Panel B1) were underpowered; I agreed that they were. That is not the same analysis as pooling several comparisons, using fixed effects, to form the 1998-and-1999-only analysis I provided a week later. JSG did not ask about power in that new analysis, and instead incorrectly inferred that my response to their questions about Figure 1 applied to the new table.

JSG's other comments on my work seem to reflect a desire to apply the research and reporting norms of medical studies to social science research. For example, JSG write, *"Important findings of no apparent effect on height and height-for-age were not reported in the abstract until the 2016 version."* This is true. The norm in economics is to strictly limit the length of abstracts; as a result, abstracts typically report only the main results. Readers may infer, from the language used by JSG, that the earlier versions of my paper purposely obscured null results on height and height-for-age. This inference, which I am sure was not intended, would be incorrect. In its 2015 and 2014 versions, non-effects on height and height-for-age were presented in both the paper's introduction and its first results table. Moreover, the content of the abstract is not recommended as an object of scrutiny anywhere in the Cochrane bias assessment tool; it arises through a modification of the Cochrane bias assessment presented as an original contribution of JSG. However interdisciplinary our standards are, one cannot reasonably expect papers to adhere to guidelines that are only announced after the fact.

Moving the goalposts in this way needlessly muddies the waters of JSG's work. This theme is reflected in the earlier discussion of statistical power as well: the Cochrane tool for assessing bias (8.15.2) explicitly recommends *against* the discussion of statistical power in appraisals.[6] The criticisms of Ozier 2016 thus appear factually without basis, while also failing to adhere, procedurally, to the guidance provided by the Cochrane bias assessment tool. There are other misunderstandings in relation to Cochrane criteria, such as in the discussion of Croke 2014, in which JSG appear to conflate "attrition" caused by random sampling with systematic "attrition bias" as the Cochrane handbook (8.4.4) describes it.[7] In Croke 2014, not all original study participants were sampled, but this was by design.

JSG's central critiques of Ozier 2016 are predicated on misunderstandings of fixed effects; of statistical power; and of the formatting of abstracts. These critiques are misguided; JSG's conclusion—that there are "substantial problems" in Ozier 2016—is incorrect.

Some other critiques that JSG present are entirely accurate, if one accepts the premise that social science should adhere to medical trial protocols. A striking phrase appears in JSG's abstract: *"None of the studies used preplanned protocols nor blinded the analysis to treatment allocation."* This point appears in JSG's Table 3, in the *"Blinding of data analysis"* column, in which every study is listed as *"HIGH RISK"* because the researcher running the analysis was *"not blinded"* as to which group was treated. Here, JSG indict not just these papers, but the social sciences generally. As JSG recognize, blinding and pre-analysis plans neither were standard nor presently are standard across economics; Olken 2015 and Coffman and Niederle 2015 illustrate that recent consideration has not produced unanimous agreement on whether, for the field of economics, their benefits always exceed their costs.[8,9] Economics is a field in which most studies are not randomized trials; many studies use publicly available (often government-produced) datasets. The norms surrounding research, reporting, and publication in economics are derived from practices in the study of secondary data. Pre-analysis plans may be rendered nonsensical in the case of outcome datasets that are freely available online, as is the case for the data used by Croke 2014: how could one credibly assert, in writing a pre-analysis plan, not to have seen a dataset which anyone could easily download?

To an epidemiologist or economist who reads JSG or this response, it may seem that in our hurry to refute one another's points, we missed the opportunity to improve interdisciplinary communication on a topic where scholars from different traditions have much to contribute. This exchange raises questions that it does not answer. What description would have allowed fixed effects to be understood reliably across disciplines? How should blinding and pre-analysis plans work when relevant datasets are freely available online? What kind of communication could have made appraisals and commentaries like this one more constructive? Perhaps one lesson we can agree on, though it has been pointed out before, is on the importance of precise, detailed reporting of study characteristics.[10] This facilitates assessment of a study against systematic review and meta-analysis inclusion criteria, and can surely simplify the task of appraising a study.

The Cochrane handbook (8.3.4) cautions against communication with original study authors. As this case illustrates, the appraisal of work from other disciplines is rife with the possibility of misinterpretation. Even if there are circumstances where it would be preferable to adhere to the Cochrane recommendations, in my view this guidance should be amended when multiple disciplines are involved. The risks of miscommunication and misunderstanding may exceed the risks in the studies being assessed.

---

[1] Jullien S, Sinclair D, Garner P. The impact of mass deworming programmes on schooling and economic development: an appraisal of long-term studies. *Int J Epidemiol (in press)*
[2] Ozier O. *Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming.* 2016. http://economics.ozier.com/owen/papers/ozier_early_deworming_20160727.pdf (6 November 2016, date last accessed).
[3] Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomized trial: rationale, design, analysis, and reporting. 2015. *BMJ* 350:h391
[4] Miguel E, Kremer M. Worms: identifying impacts on education and health in the presence of treatment externalities. 2004. *Econometrica* 72:159–217
[5] Walker SP, Wachs TD, Grantham-McGregor S, Black MM, Nelson CA, Huffman SL, Baker-Henningham H, Chang SM, Hamadani JD, Lozoff B, Gardner JM. Inequality in early childhood: risk and protective factors for early child development. *The Lancet*. 2011 Oct 14;378(9799):1325-38.
[6] Higgins JP, Altman DG, Sterne JA. Assessing risk of bias in included studies. In: Higgins JP, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0.* The Cochrane Collaboration, 2011. http://handbook.cochrane.org/ (6 November 2016, date last accessed).
[7] Croke K. *The Long Run Effects of Early Childhood Deworming on Literacy and Numeracy: Evidence from Uganda.* 2014. http://scholar.harvard.edu/files/kcroke/files/ug_lr_deworming_071714.pdf (6 November 2016, date last accessed).
[8] Olken B. Promises and Perils of Pre-Analysis Plans. *J Economic Perspectives.* 2015;29(3):61-80
[9] Coffman LC, Niederle M. Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible. *J Economic Perspectives.* 2015;29(3):81-98
[10] Evans DK, Popova A. What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. 2016. *World Bank Res Obs* 31(2):242-270