International Journal of Epidemiology, 2017, 1–4
doi: 10.1093/ije/dyw350
Commentary

Commentary

Commentary: Assessing long-run deworming impacts on education and economic outcomes: a comment on Jullien, Sinclair and Garner (2016)

Sarah Baird,[1] Joan Hamory Hicks,[2] Michael Kremer[3] and Edward Miguel[4]*

[1]Milken Institute School of Public Health, George Washington University, Washington, DC, USA, [2]University of California, Center for Effective Global Action, Berkeley, CA, USA, [3]Department of Economics, Harvard University and NBER, Cambridge, MA, USA and [4]Department of Economics, University of California, Berkeley, and NBER, Berkeley, CA, USA

*Corresponding author. Department of Economics, 508-1 Evans Hall #3880, University of California, Berkeley, CA 94720, USA. E-mail: emiguel@berkeley.edu

Accepted 11 November 2016

Introduction

Jullien, Sinclair and Garner (2016)[1] (henceforth JSG) state that they seek to 'appraise the methods' of three recent papers that estimate long-run impacts of mass deworming on educational or economic outcomes. This commentary focuses on their discussion of Baird, Hicks, Kremer and Miguel (2016)[2] (henceforth Baird). We welcome scrutiny of our work, and appreciate the opportunity to discuss JSG.[1]

Baird[2] finds evidence of gains in some educational and labour outcomes 10 years after a deworming programme in 75 Kenyan primary schools. Some gains are found in the full sample, and others among either males or females, in ways that are sensible given the context, e.g. there are gains in manufacturing employment among males but not females, fewer of whom work in this sector in Kenya.

Below we discuss JSG's claim that the evidence in Baird[2] is unreliable. It is not surprising that any two scholars might interpret a body of results differently, but JSG[1] make a series of claims that appear overstated or are somewhat misleading. Due to word limits, we discuss some points here and others in the Supplementary Appendix (available as Supplementary data at *IJE* online).

Discussion of JSG

JSG[1] do not make a substantive critique that the results in Baird are inaccurate or not robust to alternative specifications, presumably because they did not identify such issues. Rather, they make a methodological critique, namely that the results in Baird[2] are unreliable due to potentially selective reporting of positive results (JSG,[1] Table 3). We have several responses.

First, JSG[1] do not present any statistical evidence of selective reporting. They acknowledge that their claims are instead based on 'a narrative analysis'. Like witchcraft, it is easy to make claims about selective reporting, but difficult to prove—or disprove—whether it has occurred. In fact, most patterns presented in their tables lean against systematic reporting bias.

To start, across working paper versions of Baird[2] the same basic set of outcomes are presented with only minor adjustments (typically in response to suggestions by colleagues or journal referees). Furthermore, many of the robustness checks in Baird[2]—examining alternative outcomes or statistical specifications, and multiple testing adjustments-are included precisely to address reporting concerns. If the main results reported in Baird[2] were simply

false-positives, then perhaps only roughly 5% of all tests in Baird[2] and its lengthy appendix would be significant at 95% confidence, but the proportion is an order of magnitude higher. If we had attempted to 'cherry-pick' results, the proportion of significant results should have risen across versions of Baird, but it remains entirely stable (see JSG,[1] Appendix 3).

JSG[1] focus on the fact that Baird[2] present results for the full sample as well as by gender. Since the gender breakdown was not present in the first, incomplete, 2011 working paper version, JSG[1] imply that any discussion of gender per se constitutes evidence of selective reporting. (Note that, beyond gender, there is little subgroup reporting in Baird.[2])

Yet there are ample conceptual rationales for considering impacts for women and men separately. It is standard in economics to disaggregate labour market analysis by gender (Bertrand 2011),[3] especially for young adults, given the effects of childbearing. An influential contribution, Pitt, Rosenzweig and Hassan (2012),[4] cited in Baird,[2] makes a theoretical case and presents evidence that educational impacts of health investments are likely to be larger among females, and labour impacts larger for males, in a low-income setting. Any a priori analytical plan would have specified this subgroup analysis. JSG's dismissal of the results for females trivializes the importance of gender in low-income countries like Kenya, where women and men face starkly different economic opportunities.

Another aspect of the selective reporting discussion relates to the multiple testing adjustments in Baird.[2] In contrast to Baird,[2] JSG[1] claim (JSG,[1] Table 6) that the main results are not robust to multiple testing correction. We found this discussion to be among the least informative parts of JSG.[1] The data in the 'Effect robust to adjustment for multiple inference?' column ignores the fact that the adjusted *P*-values corresponding to the 'No' values range from 0.07 to 0.13, in other words near traditional significance levels even after adjustment (see our Table 1). Baird[2] report these values but JSG[1] opt not to mention them. By overemphasizing small changes in *P*-values around the arbitrary 0.05 threshold, JSG[1] create the impression that results are fragile when that is not the case. Moreover, deworming effects on the most comprehensive measure of living standards in Baird,[2] the meals eaten outcome, remain significant at the 0.05 level even after adjustment.

The final column in JSG's Table 6 ('Effect consistent across related outcomes?') is also largely uninformative. The goal of the multiple testing adjustment is to account for a set of results; cherry-picking one outcome in a broader family that is not significant and highlighting it as evidence of a lack of robustness, as JSG[1] do, is less scientific. For example, for the Baird[2] finding that males who received more deworming work more in manufacturing jobs, JSG[1] argue that there are no related outcomes with statistically significant effects; however, in fact these men also have significantly higher labour earnings.

JSG[1] also critique approaches to the presentation of results in Baird[2] that are standard in economics and other social sciences, but do not conform to norms in their own field. For instance, JSG[1] mention—at least a dozen times! —whether or not results are reported in Baird's abstract, and emphasize that the height results are not reported there. Yet it is not surprising that this result did not make it into Baird's abstract: the structure of economics abstracts is not standardized, unlike public health articles, and they are short, typically 100–250 words (the Baird[2] abstract has 146 words). Instead, economics articles usually summarize results in the introduction. The height results in Baird[2] are reported in the introduction, as well as in the main text and tables.

JSG[1] also emphasize that Baird[2] lack a pre-analysis plan. Whereas it is true that Baird[2] did not register a pre-analysis plan, such plans were until recently largely unknown in economics, and the American Economic Association RCT registry was only established in 2013.

At times, JSG[1] appeal to Cochrane review results (Taylor-Robinson *et al.* 2015)[5] to bolster their case. However, the Cochrane review results are problematic (Montresor *et al.* 2016),[6] with an incomplete sample of studies, improper selective exclusion of a study that shows weight gains (e.g. Stephenson *et al.* 1993),[7] and an underpowered statistical test. Croke *et al.* (2016)[8] show that mass deworming leads to child weight gains at the community level.

Conclusion

The issue of selective reporting raised by JSG[1] is potentially important, but the evidence presented in JSG[1] does not change our interpretation of Baird[2]. It is JSG's right to interpret the evidence in their own way, of course, but we cannot help but feel that a more even-handed discussion would have been more productive for scientific progress. A more scientific assessment would discuss Baird's strengths as well as weaknesses, for example: the value of its long-term longitudinal data, which allow estimation of the benefit-cost ratio for mass deworming and suggest that long-run income gains might be 100 times the (small) initial cost. A more even-handed appraisal would not cherry-pick null results to highlight, present multiple testing adjustments in a tendentious fashion (in JSG[1] Table 6) or summarily dismiss analysis by gender in this context. The discussion could have mentioned a methodological strength of Baird,[2] namely, the fact that two orthogonal

Table 1 Unadjusted *P*-values and adjusted q-values, added to Jullien *et al.* Table 6 (2016)[1]

Outcomes reported in the abstract		Effect robust to adjustment for multiple inference?		
		Unadjusted *P*-value	JSG claim	Adjusted q-value
Men	'stay enrolled for more years of primary school'	0.022	No	0.071
	'work 17% more hours each week'	0.017	No	0.083
	'spend more time in non-agricultural self-employment'	0.066	Remains borderline	0.133
	'spend more time in manufacturing'	0.015	No	0.083
	'miss one fewer meal per week'	0.003	Yes	0.031
Women	'one quarter more likely to have attended secondary school'	0.022	No	0.084
	'reallocated time from traditional agriculture into cash crops'	0.031	No	0.103
	'reallocate time from traditional agriculture into non-agricultural self-employment'	0.025	No	0.103

All but the last column reproduced from Jullien *et al.* Table 6 (2016).[1] The last column includes the adjusted false discovery rate (FDR) q-values from the Supplementary Appendix of Baird *et al.* (2016).[2]

sources of variation—a cost-sharing experiment carried out in a random subset of schools (which lowered deworming drug take-up), and variation in exposure to cross-school treatment spillovers—both reinforce the main results.

It may be worth stepping back and thinking about the broader public policy debate regarding deworming. The decision to fund mass deworming should be based on comparing its expected costs and benefits, so even a small probability that the effects in Baird[2] are present would make the cost effectiveness analysis favourable. To be very concrete, was the Indian Government's recent decision to carry out mass school-based deworming—at pennies per dose (using safe and approved drugs) in areas with widespread infections—misguided and not 'informed by robust evidence', as JSG[1] suggest? It appears that even JSG[1] agree that deworming might be sensible and cost-effective in such a setting, when they write: 'If a community in a given setting has a high prevalence of untreated worm infections, then mass-deworming programmes may well be an effective way to reach and treat a large number of children'.

The long-run benefits found in Baird,[2] Croke (2014)[9] and Ozier (2016),[10] as well as Bleakley (2007),[11] medium-run schooling impacts reported in Miguel and Kremer (2004)[12] and the short-run child weight gains in Croke *et al.* (2016),[8] taken together may lead mass deworming to go from being merely 'very cost-effective' to 'extremely cost-effective'. Either way, the logic of mass drug deworming administration in endemic regions appears as clear today as it was when the World Health Organization began supporting this policy decades ago.

Supplementary Data

Supplementary data are available on the *IJE* website.

Acknowledgements

We thank Kevin Croke and Owen Ozier for useful discussions. All errors remain our own.

Conflict of interest: M.K. is a former board member of Deworm the World, a US non-profit organization. He has received no funding from Deworm the World. He is also a part-time employee of USAID, which financially supports deworming activities, among its many other activities. This paper was written in his academic capacity, and neither Deworm the World nor USAID had any influence over the writing of this paper.

References

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* 2016; in press.
2. Baird S, Hicks JH, Kremer M, Miguel E. Worms at work: Long-run impacts of a child health investment. *Q J Econ* 2016;**131:** 1637–80.
3. Bertrand M. New perspectives on gender. In:. Ashenfelter O, Card D (eds). *Handbook of Labor Economics* Amsterdam: Elsevier, 2011.
4. Pitt MM, Rosenzweig MR, Hassan MN. Human capital investment and the gender division of labor in a brawn-based economy. *Am Econ Rev* 2012;**102:**3531–60.
5. Taylor-Robinson D, Maayan N, Soares-Weiser K, Donegan S, Garner P. Deworming drugs for soil transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin, and school performance. *Cochrane Database Syst Rev* 2015;**7:**CD000371.
6. Montresor A, Addiss D, Albonico M *et al*. Methodological bias can lead the Cochrane Collaboration to irrelevance in public health decision-making. *PLoS Negl Trop Dis* 2015;**9:** e0004165.
7. Stephenson LS, Latham MC, Adams EJ, Kinoti SN, Pertet A. Physical fitness, growth and appetite of Kenyan School Boys with Hookworm, Trichuris trichiura and Ascaris lumbricoides infections are improved four months after a single dose of albendazole. *J Nutr* 1993;**123:**103646.
8. Croke K, Hicks JH, Hsu E, Kremer M, Miguel E. *Does Mass Deworming Affect Child Nutrition? Meta-analysis,*

Cost-effectiveness, and Statistical Power. Working Paper #22382. Washington, DC: National Bureau of Economic Research, 2016.

9. 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 (4 December 2016, date last accessed).
10. 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 (4 December 2016, date last accessed).
11. Bleakley H. Disease and development: evidence from hookworm eradication in the American South. *Q J Econ* 2007; **122**:73–117.
12. Miguel E, Kremer M. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 2004;**72:**159–217.