Exploiting randomized exposure to early childhood deworming programs to study long run effects: A research program in progress

Kevin Croke

Development Economics Research Group, Impact Evaluation Team, The World Bank Corresponding author. 1818 H Street NW, Washington, DC 20433, USA. Email: kcroke@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 (February 6, 2017); the final version may be accessed online here:

A recent article by Jullien et al. [1] offers a critical appraisal of three studies which address the long run impact of deworming, including my 2014 working paper, "The Long Run Effects of Early Childhood Deworming on Literacy and Numeracy: Evidence from Uganda." [2] This working paper is part of an ongoing research project, which continues to incorporate new data from the study setting as it is publicly released. First I offer responses to the authors' critique of the 2014 working paper, and then provide information about the ongoing research project.

My 2014 working paper examined the long run impact of deworming, delivered through an earlier cluster randomized deworming trial by Alderman et al. [3], on numeracy and literacy outcomes. In the original trial, all children (treatment and control) received a set of basic health services such as Vitamin A supplementation and vaccination, while those in the treatment group were given these interventions plus deworming treatment. Interventions were delivered every 6 months over a period of three years. Examination of the long run impact of these interventions required a measurement of educational outcomes, for which I used a large-scale survey known as Uwezo, which since 2010 has measured basic numeracy and literacy annually in Uganda, Kenya, and Tanzania. The geographic comprehensiveness of the Uwezo survey meant that literacy and numeracy data from the 2010 and 2011 survey rounds was available for a random sample of 710 children from the program's target age group during the program period (from 2000 to 2003) in 22 out of the original 50 trial communities.

The authors' primary critique is that my study uses this subset of the clusters from the original deworming trial; the authors state that as a result, "the analysis is at high risk of attrition bias, due to loss of clusters." I disagree that there is *a priori* reason to expect bias in the estimated treatment effect. The communities for the original trial were partitioned into treatment and control randomly (via coin toss), and Uwezo documents state that their survey's selection of communities was also random within every district of Uganda, and that the selection of households was by systematic random sampling within villages [4-6]. Thus the clusters sampled by Uwezo should be, by design, a random subset of the full set of trial communities. A subsample of the original trial clusters that is randomly

generated in this way should not be a biased sample.

Selective or biased sampling should also manifest as imbalance between treatment and control. While imbalance is possible given the relatively modest number of clusters, in my paper I found no significant differences between treatment and control communities across a range of demographic and socioeconomic variables [2]. While the authors state in table 3 that "some confounders (access to water and private education) appear unbalanced," both the balance table provided in the original working paper and an augmented version, which I shared with the authors upon their request, show no significant differences between treatment and control communities on these variables. The access to water variable, which the authors highlight, is not significantly different between treatment and control despite relatively large differences because of a very high intra-cluster correlation for this variable (0.72) within communities.

This point relates closely to another criticism, which is that the analysis does not address migration. Since the survey captures a random selection of treatment and control communities, there is no a priori reason to expect that selection into treatment 10-11 years earlier would affect migration patterns in late childhood or early adolescence. Nonetheless, migration could be an issue in two ways, both of which work against the possibility of detecting a significant effect. First, it adds measurement error to the definition of treatment, attenuating estimated treatment effects [7]. Second, while preschool and primary school age children are unlikely to migrate for school or work, post-primary school aged adolescents (the 14, 15, and 16 year olds in the Uwezo sample) might migrate for these reasons. But the only reason they would migrate *differentially* in the treatment versus control group would be if in fact the treatment had affected human capital in some way. If literacy and numeracy are positively associated with migration (such as to attend secondary school or pursue work in urban areas), than any positive treatment effect will be underestimated.

The authors also argue that the study "is substantially underpowered" to assess long run effects by comparing it to the sample in Ozier [8]. Given the low cost of deworming, I also would greatly prefer increased statistical power, since even small benefits would likely make deworming a cost effective intervention. However it is unclear why a comparison with Ozier's sample size is the relevant metric. That study estimated benefits which accrued to children via epidemiological spillovers, while the focus of my paper is on the direct deworming of children. Direct treatment should have larger effects, thus requiring less statistical power. Also with respect to power, the authors note that statistically significant results (at $p<0.05$) are only present when covariates such as age, gender, and survey round are included. Setting aside the fact that the unadjusted p-value for numeracy is quite close to the 0.05 threshold ($p=0.065$), specifications with pre-treatment covariates are equally unbiased but are more efficient, and therefore help maximize statistical power [7]. Moreover, age and gender are standard covariates in the education literature, and are important predictors of literacy and numeracy. This is demonstrated by the difference in R^2 between unadjusted models, where it ranges from 0.01 to 0.03, and fully adjusted models, where it is between 0.34 and 0.39.

Taking a step back from these specific points, I do agree with the authors that having more data would be better. This project originated as an attempt to use existing data to shed light on a question that is currently poorly understood, by leveraging pre-existing experimental variation in access to deworming and an unrelated data collection project. The downside of using existing surveys is that one does not control the sampling strategy. This is an ongoing challenge in research on the long run impact of development interventions.

The good news, however, with respect to both statistical power and the comprehensiveness of the sample is that more data is being collected over time. Since 2010, Uwezo has conducted annual surveys to test basic literacy and numeracy in Uganda, randomly sampling 30 communities in every district. Each year, 20 villages are retained from the previous year's sample and 10 are newly selected. Thus, the study sample will gradually grow over time as Uwezo progressively samples more of the original trial communities. New data will enhance coverage of the original sample clusters, although with the cost that more of the program-exposed cohorts will have reached ages where out-migration is more likely.

Awareness that data would accumulate over time has guided our research process to date. In economics it is common to circulate working papers with preliminary results, which are updated and revised based both on comments from readers, and in light of new data. Thus, after the initial working paper was publicly posted in 2014, we waited for the release of an additional year's data (from 2012) before revising the paper for submission. (Prof. Rifat Atun joined the research team at this stage). Following medical journal conventions, this version was not posted publicly as a working paper, but the main finding of positive treatment effects was strengthened by the addition of the 2012 data. Additional data which can also be matched to the original study communities from the 2013-2015 Uwezo survey rounds is likely to be publicly available in the near future. A revised version of the manuscript will be prepared when this data is complete and available.

Jullien et al. [1] also note that this paper did not have a pre-analysis plan. While pre-analysis plans are now common for RCTs in economics, they are relatively unusual for work which relies on secondary or public data sources. In such cases it is more typical to present extensive robustness checks, in order to demonstrate how results do or do not change as functional form (for example) or choice of covariates is varied. We also took the further step of sharing data and code with external researchers to test the robustness of our results. In 2014 I shared pre-publication data and programs with researchers at Givewell, an organization which evaluates charities, including those that implement deworming programs. These researchers replicated the original (2014) working paper, and in 2015 I shared data and code for a revised version of the paper, which included the 2012 data. This follows Olken's [9] suggestion that data sharing and replication are an alternative to pre-analysis plans as a way to ensure robustness of findings.

To conclude, while I disagree with the critiques of Jullien et al [1] with respect to my 2014 working paper for the reasons stated above, I also note that new results should be expected as additional rounds of data become fully available from Uwezo.

References:

1. Jullien S, Sinclair D, Garner P. "The Impact of Mass Deworming Programs on Schooling and Economic Development: An Appraisal of Long-term Studies." 2016. Int J Epi, this issue.
2. Croke K. "The Long Run Effects of Early Childhood Deworming on Literacy and Numeracy: Evidence from Uganda." 2014. Working paper, available at http://scholar.harvard.edu/files/kcroke/files/ug_lr_deworming_071714.pdf. Last accessed 4 November 2016.
3. Alderman H, Konde-Lule J, Sebuliba I, Bundy D, Hall A. "Effect on weight gain of routinely giving albendazole to preschool children during child health days in Uganda: cluster randomised trial. 2006. Br Med J 333(7559): 122.
4. Uwezo Uganda. "Are Our Children Learning? Annual Learning Assessment Report." 2010. Kampala, Uganda: Uwezo Uganda. Available at http://www.uwezo.net/wp-content/uploads/2012/08/UG_2010_AnnualAssessmentReport1.pdf Last accessed 3 November 2016.
5. Uwezo Uganda. "Are Our Children Learning? Annual Learning Assessment Report." 2011. Kampala, Uganda: Uwezo Uganda. Available at http://www.uwezo.net/wp-content/uploads/2012/08/UG_2011_AnnualAssessmentReport.pdf Last accessed 3 November 2016.
6. Uwezo Uganda. Uwezo 2011 Assessment Sampling Report. Kampala, Uganda: Uwezo Uganda.
7. Angrist JD, Pischke JS. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton US and Oxford UK: Princeton University Press, 2009.
8. Ozier O. Exploiting Externalities to Estimate the Long-term Benefits of Early Childhood Deworming." 2016. Working paper, available at http://economics.ozier.com/owen/papers/ozier_early_deworming_20160727.pdf. Last accessed November 2 2016.
9. Olken B. Promises and Perils of Pre-Analysis Plans. 2015. J Econ Perspectives, 29(3): 61-80.