adequately powered analysis methods (Figure 2). These results contribute to a growing literature using cluster randomized or quasi-experimental designs to study deworming's socioeconomic impacts, all of which estimate positive long-run impacts on educational and labour market outcomes (Ahuja *et al.*,[9] Bleakley,[10] and Ozier[11]).

Supplementary Data

Supplementary data are available at *IJE* online.

Acknowledgements

We thank Kevin Audi, Evan DeFilippis, Felipe Gonzalez, Leah Luben and especially Michael Walker for excellent research assistance. All errors remain our own.

Conflict of interest: MK is Scientific Director for Development Innovations Ventures at USAID. Among its many other activities, USAID supports deworming.

References

1. Aiken A, Davey C, Hargreaves J, Hayes R. Re-analysis of health and educational impacts of a school-based deworming program in western Kenya: a pure replication. *Int J Epidemiol* 2015; **44:**1572–80.
2. Davey C, Aiken A, Hayes R, Hargreaves J. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial., *Int J Epidemiol* 2015; **44:**1581–92.
3. Miguel E, Kremer M. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 2004;**72:**159–217.
4. Aiken A, Davey C, Hargreaves J, Hayes R. *Re-Analysis of Health and Educational Impacts of a School-Based Deworming Program in Western Kenya: Part 1: Pure Replication*. 3ie Replication Paper 3, part 1. Washington, DC: International Initiative for Impact Evaluation (3ie), 2014.
5. Hicks JH, Kremer M, Miguel E. (a) *Estimating Deworming School Participation Impacts and Externalities in Kenya: A Comment on Aiken et al*. Original author response to 3ie Replication Paper 3, Part 1. Washington, DC: International Initiative For Impact Evaluation (3ie), 2014.
6. Davey C, Aiken A, Hayes R, Hargreaves,J. *Re-Analysis of Health and Educational Impacts of a School-Based Deworming Program in Western Kenya: Part 2: Alternative Analyses*. 3ie Replication Paper 3, Part 2. Washington, DC: International Initiative For Impact Evaluation (3ie), 2014.
7. Hicks JH, Kremer M, Edward M. (b). *Estimating Deworming School Participation Impacts in Kenya: A Comment on Aiken et al*. Original author response to 3ie Replication Paper 3, part 2. Washington, DC: International Initiative for Impact Evaluation (3ie), 2014.
8. Aiken A, Davey C, Hayes R, Hargreaves J. *Deworming Schoolchildren in Kenya – Replication Plan*. Washington, DC: International Institute Impact Evaluation (3ie), 2013.
9. Am Ahuja, Baird S, Hamory Hicks J, Kremer M, Miguel E, Powers S. *When Should Governments Subsidize Health? The Case of Mass Deworming*. National Bureau of Economic Research Working Paper #21148. Cambridge, MA: National Bureau of Economic Research, 2015.
10. Bleakley H. Disease and development: evidence from hookworm eradication in the American South. *Q J Econ* 2007;**122:**73–117.
11. Ozier O. *Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming*. World Bank Policy Research Working Paper #7052, Washington, DC: World Bank, 2014.

Authors' Response to: Deworming externalities and school impacts in Kenya

International Journal of Epidemiology, 2015, 1596–1599
doi: 10.1093/ije/dyv130
Advance Access Publication Date: 22 July 2015

James R Hargreaves,[1] Alexander M Aiken,[1]* Calum Davey[1] and Richard J Hayes[2]

[1]Department of Social and Environmental Health Research, and [2]Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK

*Corresponding author. E-mail: alexander.aiken@lshtm.ac.uk

We thank Hicks, Kremer and Miguel (hereafter HKM) for their responses to our replication analyses[1] of Miguel and Kremer's 2004 study (hereafter M&K).[2] Here, we reflect on the background to this work and our conclusions, respond to two core criticisms and offer some concluding thoughts.

Background

Our interest in replicating M&K came not from an interest in school-based deworming but in evaluation methodology. During the early 2000s, economist-led randomized trials emerged on the effects of cash or in-kind incentives on HIV-risk behaviours.[3] As HIV epidemiologists, we found appraising these studies challenging because of different approaches to study design, reporting and analysis. We replicated M&K to learn more about the techniques used in economist-led trials, by re-analysing and presenting this famous trial in a format familiar to epidemiologists.

Let us be clear: we have not reviewed the evidence base on the effects of school-based deworming and make no policy recommendations in this area. We did not set out to disprove the findings of the original study: our analysis and inference have been guided only by our experience in cluster-randomized trials. We have great admiration for the original authors, whom we have not set out to denigrate (unhelpful attempts to characterize these exchanges as 'worm wars', notwithstanding[4]). Readers would do well to consider how their own work would stand up to the level of scrutiny to which we have subjected M&K, who have been open with us throughout.

Our conclusions

In relation to the 'externalities' investigated in M&K, our conclusions are limited to 'pure replication' of the results presented in the original paper and using the original authors' analytical methods.[5] Most results were replicated. However, our re-analysis changes an important conclusion of the original paper: after correction of coding errors, we find little evidence for indirect effects of the intervention over a distance of 0–6 km from treatment schools.

Our 'statistical replication'[6] finds that the trial provides some evidence for a positive effect of the combined education and deworming intervention on school attendance among children eligible for deworming. The strength of evidence is sensitive to analytical choices and is weak in our preferred, conservative approach. Consistent with M&K, we find no evidence of an effect on examination performance.

Schools	Year 1 (1998)	Year 2 (1999)
Group 1 (n=25)	Intervention	Intervention
Group 2 (n=25)	Control	Intervention
Group 3 (n=25)	Control	Control

Figure 1. Diagram of the roll-out of the intervention, adapted from the pre-analysis plan.

We are concerned about the risk of bias in the trial for two main reasons. First,the lack of an available pre-defined sampling strategy for school visits to assess attendance became evident as we analysed the trial profile.[7] In further investigation, we noted unexpected patterns: the amount of outcome data differed by arm and study year, and was associated with level of school attendance.[6] Moreover this association appeared to differ by study arm, with oversampling of schools with high attendance in the intervention arm but under-sampling in the control arm. We have been unable to find a clear explanation for these patterns, but could not rule out the possibility that effects on school attendance may have been biased as a result.

Second, we are concerned about the potential for bias in the intervention effect arising from year 1 to year 2 changes in school attendance in this stepped-wedge trial design. Both the analysis model we originally proposed and that of the original authors produce an intervention effect that is influenced strongly by changes in Group 2 from year 1 to year 2 as that group crosses over from control to intervention status. This is a non-randomized, before-and-after comparison where there is a possibility of underlying secular change that may not be accounted for adequately by the statistical model. The strict comparison between randomized groups within each year yields effects that are substantially smaller in magnitude than the pooled effect.

Response to criticisms

HKM consider that we report externalities at a distance that means that the estimates are 'too noisy to be informative'. We understand their original externality analyses were exploratory and we value their innovative thinking in this regard. We simply identified and corrected errors in the original tables, leading us to conclude that there is little evidence, under the authors' original analysis specification, for externalities at the distance (0–6 km) over which these were estimated in M&K.[2] HKM do not dispute this conclusion, but proceed to report further analyses exploring externalities over a range of distances, and contend that we should have followed this approach. On this we must disagree, while hoping that appraisers of the evidence will consider their further analyses. Our view is simply that these were, at the time, innovative hypothesis-generating analyses worthy of further consideration. As with all exploratory analyses, we caution against over-interpretation of effects seen at specific distances defined after examination of the data.

HKM describe as an 'error' our decision to include data on school attendance in each year before administration of deworming, saying this decision is unjustified and not in line with our published analysis plan.[8] We disagree.

Figure 1 in our analysis plan indicates our understanding that the cross-over point for schools from control to intervention was in line with calendar year. From the outset, we understood the intervention package (comprising both health education and deworming medications) to have been delivered across full school years, with effects on school attendance evaluated through assessments in that same school year. Nevertheless, when concerns about our approach were voiced, we carried out additional sensitivity analyses which treat the outcome data differently, one of which is in line with the approach used in M&K.[2] Our conclusions are unchanged. There is statistical evidence for an effect on school attendance in some but not all analyses, and we consider the evidence from the trial to be at risk of bias.

Final thoughts

We have two recommendations. First, that an updated synthesis of the evidence on school-based deworming is undertaken: we hope both our findings and the responses of HKM will be taken into account. Second, that this body of work should prompt discussion between epidemiologists and economists in the impact evaluation community of the benefits that would be gained from greater standardization in the way we design, analyse and report studies.

Replication in this form is new to us as epidemiologists. It has great potential: re-appraisal of influential papers such as M&K must surely have value for the scientific community and policy makers. However, much work remains to ensure this promise is fulfilled. The importance of pre-defining aspects of the research process is an area of contention that runs through the discussion. The absence of a pre-defined sampling strategy for school attendance assessment is a driver of our assessment that the trial findings should be considered as at risk of bias. HKM suggest that we deviated from our pre-defined analysis plan. We strongly support the importance of pre-specified plans for data collection and analysis. But in this replication work, where we did not design or implement the primary data collection or intervention activities, at times we felt that the most appropriate course of action needed to change.We came to conclude that the final model that we pre-specified was not optimal to reflect the conduct and results of the study—although it was reported fully and as the primary analysis in our paper.[6] Managing this tension between pre-specification and adaptation will be an important challenge for future replications.We also suggest that future replication studies will need higher funding and stronger processes for mediating the inevitable dynamics between original and replication researchers.

There is much to be gained from researchers from different disciplines engaging with the same problems. But this is not always easy. As one example from this work, we agree with HKM that in the presence of spill-over, or externality, effects, a simple comparison between individuals or schools randomized to intervention and control conditions will be biased toward the null. Epidemiologists generally describe this phenomenon as 'contamination'. We see potential contamination as a limitation of the original trial, one that might have been mitigated through different design choices even if externalities were the primary effect of interest. In contrast, we perceive that our economist colleagues see the exploration of spill-overs as a great strength of M&K. The impact evaluation movement must surely accommodate this full range of perspectives: something will be lost if we cannot communicate across the disciplinary divide. M&K's work was undertaken over 15 years ago, and they have subsequently been at the forefront of the movement towards rigorous use of randomized field experiments in development economics.[9] We have much to learn from them and much to share. Some standardization of practice in the design, analysis and reporting of impact evaluation studies across disciplines would, we feel, not constrain but rather enable the combined strengths of our disciplines to be brought to bear on the world's health and development challenges.

References

1. Hamory Hicks J, Kremer M, Miguel E. Deworming externalities and schooling impacts in Kenya: A comment on Aiken *et al.* (2015) and Davey *et al.* (2015). *Int J Epidemiol* 2015;**44:** 1593–96.
2. Miguel E, Kremer M. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 2004;**72:**159–217.
3. Pettifor A, MacPhail C, Nguyen N, Rosenberg M. Can money prevent the spread of HIV? A review of cash payments for HIV prevention. *AIDS Behav* 2012;**16:**1729–38.
4. Özler B. *Worm Wars: a Review of the reanalysis of Miguel and Kremer's Deworming Study. Development Impact*. Washington, DC: World Bank, 2015.
5. Aiken AM, Davey C, Hargreaves J, Hayes RJ. Re-analysis of health and educational impacts of a school-based deworming program in western Kenya: a pure replication. *Int J Epidemiol* 2015; **44:**1572–80.
6. Davey C, Aiken AM, Hayes RJ, Hargreaves J. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial. *Int J Epidemiol* 2015;**44:** 1581–92.
7. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trials* 2010;**11:**32.

8. Aiken AM, Davey C, Hayes RJ, Hargreaves J. *Deworming Schoolchildren in Kenya – Replication Plan*. Washington, DC: International Institute Impact Evaluation (3ie), 2013.

9. Duflo E, Kremer M. Use of randomization in the evaluation of development effectiveness. In: *Evaluating Development Effectiveness*. Washington, DC: World Bank, 2005.

Commentary: Replication of influential trial helps international policy

International Journal of Epidemiology, 2015, 1599–1601
doi: 10.1093/ije/dyv131
Advance Access Publication Date: 22 July 2015

Paul Garner,[1] David Taylor-Robinson[2] and Harshpal Singh Sachdev[3]*

[1]Evidence Synthesis for Global Health Group, Liverpool School of Tropical Medicine, Liverpool, UK, [2]Department of Public Health and Policy, University of Liverpool, Liverpool, UK and [3]Sitaram Bhartia Institute of Science and Research, New Delhi, India

*Corresponding author. Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA GB. E-mail: Paul.Garner@lstmed.ac.uk

Accepted 8 June 2015

The World Bank, the Gates Foundation and the World Health Organization promote national child deworming programmes in developing countries.[1] They assert these programmes will improve nutritional status, health and school performance, and hence contribute to economic growth. Indeed, the World Health Organization states that deworming contributed to Japan's economic boom in the 1950s,[2] and Nobel Laureates meeting in Copenhagen ranked deworming as the fourth most important intervention to solve the health problems of the whole world.[3]

Surprisingly, the evidence base for these claims from controlled studies is limited. Critically, according to the Cochrane review which two of us author, there is quite good evidence of no effect for the main biomedical outcomes in deworming, making the broader societal benefits on economic development barely credible (Figure 1).[4] Nevertheless, the advocates increasingly rely on a single large quasi-randomized trial carried out in Kenya, published in 2004 in *Econometrica*,[5] which reports school attendance.

This study has been highly influential. The International Initiative for Impact Evaluation commissioned the London School of Hygiene and Tropical Medicine (LSHTM) to replicate the analysis, as the original analysis is 'based on econometric approaches and used a language and format that would be unfamiliar to many health care researchers'.[6] The replication aimed to provide detail of the methods and reporting in line with the CONSORT statement. The team are internationally recognized, independent and meticulous in their approach. They agreed a protocol, carefully checked and corrected the raw data, and then re-ran their prespecified analysis.[6]

Their first paper is a pure replication,[7] exactly repeating the authors' original analysis. This paper clarifies some methodological details not provided in the original paper, but it also uncovers a series of important coding and analysis errors. Some of the corrected results are consistent with the original findings, but others are quite different. Most notably, the much quoted 'positive externalities'—where the benefits of treating children in one school 'spill over' to benefit children in adjacent schools—vanish in their corrected analysis.

Their second paper uses approaches more familiar to epidemiologists, and allows a more thorough exploration of the data.[8] There are substantial amounts of missing information, and some unexpected patterns that are difficult to explain. For example, there is a correlation between the number of observations in each school and the reported attendance, with more observations associated with lower attendance reported—except in some of the intervention groups, where more frequent observation is associated with better attendance. This raises the possibility that the process of observation influenced outcome reporting and this was different in control and intervention groups.