



Original paper selected for replication:

Edward Miguel and Michael Kremer. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities (2004)." *Econometrica* 72, no. 1 (2004): 159-217.

Investigators:

Dr Alexander Aiken, Clinical Research Fellow¹
Calum Davey, Research Fellow¹

with support from

Dr James Hargreaves, Senior Lecturer^{1,2}
Prof Richard Hayes, Professor of Epidemiology and International Health¹

Affiliations:

- 1= Department of Infectious Disease Epidemiology,
Faculty of Epidemiology and Population Health,
London School of Hygiene and Tropical Medicine
- 2 = Centre on Global Health Security, Chatham House.

Glossary of terms

Epidemiological/medical term

Econometric language equivalent

indirect benefit

(positive) externality

CONSORT

Consolidated Standards of Reporting Trials

CRT

cluster randomised trial

exposure

effect from an external agent

bias

systematic error in data

contamination

spillover of effect between study groups

helminth

(parasitic) worm

1. Background

1.1. Description of original study

We are planning to replicate the analysis originally performed by Miguel and Kremer describing the impacts of a school-based deworming programme in Kenya on the health, school attendance and academic performance of school pupils (Econometrica, 2004,(1)). This paper analysed data collected as part of a school-based deworming program delivered by Internationaal Christelijk Steunfonds (ICS), a Dutch charitable organisation, to 75 schools in Busia District in western Kenya in 1998-1999. Schools were stratified by administrative area and involvement in other ICS programs and then quasi-randomised (listed alphabetically then alternating assignment) into three groups (25 school in each group, average of 400 pupils per school) and the deworming intervention was introduced in stages over several years, as shown in table 1. This phased introduction is known (in the medical literature) as a “stepped wedge” design of a cluster randomised trial – although in this study, just two “steps” were used.

| | 1998 | 1999 |
|---------|--------------|--------------|
| Group 1 | Intervention | Intervention |
| Group 2 | Control | Intervention |
| Group 3 | Control | Control |

Table 1: Stepped-wedge design of study

The intervention was composed of two elements: first, administration of anti-helminthic (deworming) treatments given in appropriate doses at spaced intervals and secondly, a package of educational interventions. Girls over the age of 12 were not intended to receive the intervention, although some did in practice. Different drug combinations were used based on the prevalence of different types of worm infections in each school prior to the intervention. Educational measures consisted of worm prevention education, including stressing the importance of handwashing, wearing shoes and not swimming in freshwater to avoid the transmission of various types of worm. The original study also mentions that there were other school-based interventions led by ICS occurring concurrently in 27 of 75 schools (p165). However, previous analyses by the study authors (2) have found that these interventions had no substantial effects, and we have therefore assumed that these had no influence on the outcomes measured in this replication.

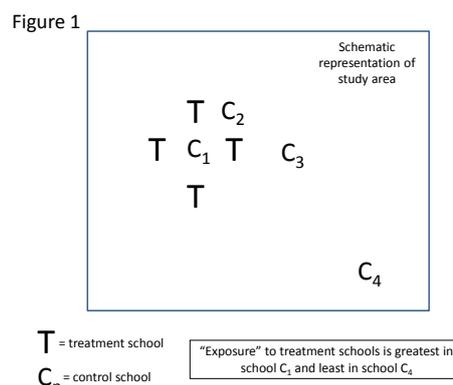
The original analysis looked at the deworming intervention to prevent four different types of worm infection: hookworm, roundworm, whipworm (all geohelminths – literally “ground worms”) and schistosomiasis. Key biological features of these infections are summarised in table 2. All schools received treatment against soil-mediated helminth infections (geohelminths), but only a subset of schools additionally received treatment against schistosomiasis, a freshwater-mediated infection. The treatment allocation was based on local prevalence of schistosomiasis infection based on parasite surveys – schools where schistosomiasis rates were found to be low were not eligible. Whilst all the schools received the treatment against geohelminths, only a minority of schools (6/25 in 1998 (Group 1 only), 16/50 in 1999 (in Groups 1+2)) received additional treatment for schistosomiasis. There is no data available from parasitic surveys to indicate which schools in Group 3 would ultimately receive treatment for schistosomiasis.

The data in this study are clustered, with schools as the unit of clustering. Analysis of cluster-randomised trials must make explicit recognition of the clustered nature of the data – in clustered data, an individual is often more likely to have a similar result to another individual within that cluster, than they are to an individual in another cluster. This has implications for how differences between groups are determined and confidence limits are estimated. The original study reported that it had accounted for clustering, but did not describe explicitly how this was performed.

The impact of the ICS program was measured in three different domains: school attendance, exam performance and health (principally worm infection, also nutritional and haematological parameters). School attendance was measured by staff from ICS performing multiple unannounced visits to all schools. Exam performance was measured in a variety of subjects in exams administered by ICS at the end of the academic year. For the health impacts, worm infection was only measured in treatment schools immediately before deworming, as it was felt unethical to test for worm infection without offering treatment. Therefore, worm infection rate was not measured in Group 2 in 1998 and in neither year in Group 3. Haemoglobin and nutritional status were only measured in randomly-

selected subsets of children. In the original analysis, an estimate of the direct benefit of the intervention in each of the three domains was obtained by comparing outcomes in schools that received the intervention to those that did not receive it in that year.

The original analysis also assessed the indirect benefits (positive externalities) of the intervention from preventing transmission of worm infections in nearby schools. This was determined using a spatial approach, illustrated schematically below (fig 1). The control schools (C_n) were at different physical proximities to treatment schools (T). As these worm infections are all transmitted by excretion of worm eggs in faeces, and as faecal contamination of the environment was known to be common, it was assumed that there would be a local reduction of transmission of worm infection in an approximately circular area up to 6km around the intervention schools, where the children attending the school were assumed to live. The authors hypothesized that the comparison schools would receive greater benefit from indirect reduction in worm infection if they were close to many intervention schools: hence (in the schematic figure) school C_1 would received a greater indirect benefit than school C_2 , in turn greater than school C_3 and so on. An additional independent term was also used in their modelling process to account for variation in local population density – schools C_1 and C_2 are in areas of greater local population density than schools C_3 and C_4 ((1), p176). The variation in indirect benefit across a gradient of exposure created by the variation in spatial proximities could then be used to estimate the overall scale of the indirect benefit. An indirect benefit within schools (untreated pupils in treatment schools) was also determined.



The original analysis concluded that there were both direct and indirect benefits to health and school attendance arising from the deworming program. They also found that there did not appear to be a benefit (either direct or indirect) of deworming on academic test scores. The findings are summarised in the table below: effects that were felt to be beneficial and significant are highlighted (sd = standard deviation, se = standard error).

| Measure | Direct effect (=treated pupils) | Indirect effect: within-school (untreated pupils in treatment school) | Indirect effect: other schools (av of pupils in control school) |
|--|--|---|---|
| School attendance (% increase) see p195-6 | +7.5% (se 2.7%) | +5.6% (se not given) | +2.0% (se1.3%) |
| Exam performance (average difference) p201 | Year 1 -0.032 sd Year 2 0.001 sd | Insignificant result (data not shown) | -0.049 sd (se 0.052) |
| Health Worm infection | -25% mod/hvy inf p173, p184 | -12% mod/hvy inf p182-4 | -23% mod/hvy inf p184-8 |
| Anaemia | Higher Hb but not significant p173 | Nil reported | Nil reported |
| Nutritional status(p173-4) | WAZ: no difference HAZ: slight benefit? | Nil reported | Nil reported |

Terminology: An overall effect refers to the total effect at the level of a cluster – this is a combination of the direct effect in treated individuals and the within-cluster indirect effect on untreated individuals within the cluster.

1.2. Impact of original study

The original study has been enormously influential in the economics literature, with a total of almost 800 citations (across all disciplines) as of July 2012. It has contributed towards deworming children being ranked by the Copenhagen Consensus Center as fourth among the sixteen most cost-effective investments to overcome the world's biggest challenges in 2012 (3). The panel recommended that US\$300 million be allocated annually for deworming, providing an endorsement to which policymakers and philanthropists are likely to look when prioritizing allocation of limited

resources. Nobel laureate economist Robert Mundell is quoted by the Center, saying: *“Deworming is an overlooked intervention deserving of greater attention and resources. This simple, cheap investment can mean a child is healthier and spends more time in school.”*

However, this paper has received much less acclaim in the medical literature, and was notably omitted from a Cochrane Review (the gold-standard of evidence-based medicine) of this subject in 2008. More recently, an updated Cochrane Review (published in May 2012) has now included this paper (4), but has described many limitations of the conduct of the original study, which could have led to bias (systematic error) in data collection. In particular, the reviewers were concerned about risk of bias from baseline imbalance, incomplete outcome data and sequence generation, and overall, the paper was graded to have “high risk of bias”. Despite these limitations, the findings arising from the analysis of these data were mentioned frequently in the review, although the reviewers ultimately concluded that *“... it is probably misleading to justify contemporary deworming programmes based on evidence of consistent benefit on nutrition, haemoglobin, school attendance or school performance as there is simply insufficient reliable information to know whether this is so.”* The debate has recently been sharpened by publication of a discussion piece titled “Deworming debunked” in the British Medical Journal (5).

The original study has unquestionably been highly influential in shaping national health policy in Kenya – in 2009, a nationwide school deworming program was implemented. Over 3.6 million children were dewormed across 8,200 schools in 2009. The Kenyan Government launched the national program, with total costs of US\$ 0.36 per child. In addition, over 1,000 district/division-level personnel and over 16,000 teachers across 45 districts were trained during program implementation. A “London Declaration” on Neglected Tropical Diseases was signed in 2012, endorsed by many agencies including the World Bank, USAID and several major pharmaceutical firms agreeing (amongst other goals) to *“Sustain, expand and extend drug access programmes to ensure the necessary supply of drugs and other interventions to help control by 2020 schistosomiasis [and] soil-transmitted helminthes”*(6).

1.3. Reasons for replication

1. This has been an enormously influential paper as described above.
2. The specific subject of the health and educational impacts of deworming programs remains of high interest in both the medical and econometric fields. As worm infections remain extremely common and deworming treatments are both cheap and highly effective, if there were proven benefits associated with deworming, this could potentially lead to huge impacts in global public health programs, especially in children in developing countries.
3. The general subject of the broader economic impacts of public health programs and how to assess these is also of great current interest. Cluster randomised trials (CRTs) are a powerful methodology for investigating such impacts, but require appropriate statistical handling to reach appropriate conclusions. However, randomised trials rarely make assessment of the benefits that accrue from spillover of the effects of an intervention – this study highlights the risks and benefits of such an analytic approach.
4. The original analysis for this study was based on econometric approaches and used a language and format that would be unfamiliar to many healthcare researchers – this may account for the limited appreciation of the study amongst “health academics” in general and epidemiologists in particular. We hope that by reframing the analysis of this data in an “epidemiological” format, this will make it accessible to a wider readership.

2. Planned replication work - Aims

The International Institute for Impact Evaluation (3ie) has agreed to fund our planned replication of this analysis, and the authors have already kindly shared their original data and analysis files with us (data received Jan-Feb 2013). Following our initial inspection of the data, we propose to conduct the following analyses.

2.1. Overview

Broadly speaking, we are aiming to conduct a pure replication¹ of the original study and then move on to a statistical replication¹ by applying an epidemiological approach for analysis of a stepped-wedge Cluster Randomised Trial (CRT) (7) to the original data. We aim to produce constructions of the two major outcomes (school attendance and exam performance) based on our own interpretations of the raw data. This will allow us to estimate the direct effect of the intervention on school attendance and exam performance, with appropriate confidence intervals. As far as possible, we aim to follow the CONSORT criteria (8) for reporting a clinical trial, including the specific adaptations relevant to cluster-randomised trials. This includes detailed descriptions of sample size determination, randomisation, blinding and encourages clear diagrams to explain the flow of participants through the study. We aim to use a “vertical” method for the analysis of the data that allows us to incorporate data from all three study groups and makes best use of the randomisation. This approach has been used previously in a stepped-wedge, cluster randomised trial (9): we feel that this makes the best use of the data in the spirit in which it was collected.

In addition, and depending on the results of the primary analyses, we will conduct further analyses that look at the direct effects on the health-related outcomes (burden of worm infections and nutritional parameters) and the indirect effects of the intervention on all three outcomes domains (school attendance, exam performance, health indicators). We aim to replicate the spatial method used in the original study to estimate the indirect effects of the intervention, using the same distances (up to 6km from schools) employed in the original study, as these are plausible distances for the scale of such an effect. However, our plan for analysis of these indirect effects is dependent on first demonstrating a direct effect – following the standard reporting practice for clinical trials, if our analysis does not demonstrate direct effects, we will not pursue analyses looking for indirect effects.

2.2. Aims of replication

The **primary aims** of this replication study are as follows:

1. To conduct a pure replication¹ of the original analysis.
To analyse in a Cluster Randomised Trial analysis format,
2. To determine if the intervention was associated with a direct effect on school attendance in treated children, as compared to untreated children of similar ages in control schools.
3. To determine if the intervention was associated with a direct effect on exam performance in treated children, as compared to untreated children of similar ages in control schools.

Both of the CRT analyses will be conducted separately for the different drug treatment types (albendazole treatment for geohelminths; praziquantel treatment for schistosomiasis).

We also aim to evaluate the following health-related outcomes as **secondary aims** of this work:

1. In a “standard” trial analysis, how much improvement was achieved, on average, on the parasite burden in the children in the treatment schools in Group 1 as compared to untreated children in Group 2. Worm burden was measured in approx 2,000 pupils with double-reader data available for the four different types of worm infection separately. We will present separate analyses for

¹ 3ie’s guidelines to replication of impact evaluations describe three types of replication: pure, statistical and scientific. Pure replication is the independent reconstruction of variables from the raw data and re-estimation using the study’s original methodologies. Pure replication should always precede the other two forms of replication. Statistical replication is the reanalysis of the study’s original hypothesis using different data treatments (eg. different variable constructs, different data sources, different data handling). Scientific replication involves the deliberate introduction of alternative conceptual causal frameworks (“theories of change”).

pupils who were reported to be compliant with the treatment and those who were reported not to have received it. We will analyse both the quantitative outcome (mean egg burden) and the binary outcome (moderate/severe infection: yes/no) for the four separate types of worm infection, using species-specific thresholds as described in table 2.

2. We aim to evaluate whether there is evidence that this program had a direct impact on nutritional status, as measured by Height-for-Age (HAZ) or Weight-for-Age (WAZ) scores. WAZ scores were collected for approx 13,000 pupils in 1998 and HAZ and WAZ were collected for approx 8,500 pupils in 1999. Again, we will report data for pupils who were reported to have received drug treatment for deworming separately from those who did not receive this. We may also attempt to calculate Weight-for-Height (WHZ) scores from raw data – this is a more useful index of the nutritional changes occurring in a short period of time.

Depending on the results of the previous aim, if there is evidence of a direct effect on improvement in nutritional status from receipt of the intervention, then we will look to see if there is evidence of an indirect effect of this intervention, as per the spatial format of the original analysis. After initial inspection of the data, and consideration of the time-frame available for this work, we will not be able to conduct the following analyses:

We are not seeking to determine whether this program had an indirect effect on the parasite burden in children in neighbouring schools (ie. an indirect effect on worm infection) as we believe that without pre-intervention data in the control schools (Groups 2 and 3, data not collected for ethical reasons), we cannot assess pre-intervention (baseline) similarity between treatment and control arms, and as such cannot conduct a worthwhile analysis. In fact, the original study reported some evidence to suggest that there may have been substantial differences in parasite burden prior to treatment between Group 1 (any moderate/heavy infection prior to treatment in 1998: 38%) and Group 2 (any moderate/heavy infection prior to treatment in 1999: 52%); Group 3 never had any parasitological testing performed (table V, p173). The authors themselves noted that there appeared to be a marked rise in worm infections between 1998 and 1999, probably attributable to flooding associated with the El Niño weather system in that year.

We are also not seeking to determine whether the school-based deworming program had a direct impact on the prevalence of anaemia (abnormally low haemoglobin concentration) as only 778 out of approximately 20,000 (~4%) of pupils had relevant testing performed, and this testing was only performed in 1999, so no baseline comparisons are possible. Although the authors have reported to us that these pupils were selected at random from children in the different study groups, we feel that there are too many other risks of bias with this Hb data to make their interpretation worthwhile. Similarly, we do not aim to analyse any self-reported or fieldworker-observed outcomes as we feel that in an unblinded study these measures are too subjective to be of practical use.

The replication will focus on the estimates of effect of the intervention on the primary endpoints as described above. In addition, we will qualitatively assess the basic cost structure assessments used by the original authors, consider how these might relate to our findings and outline the implications for further research. We will not perform extensive new work on the cost effectiveness analysis as part of this replication, in part because the programme environment in Kenya is likely to have changed since the data was collected. In contrast with the estimates of effect, we are unsure that cost effectiveness estimates from 1999 will be informative for future policy decisions.

2.3. Proposed replication plan – tackling issues raised by Cochrane Review

Obviously, this replication cannot collect further data from the original study, or change the design of the study, but we believe that this plan will address some of the important concerns expressed in the Cochrane Reviewers (4) in the following ways:

Risk of bias from incomplete outcome data: we are aiming to make a clear description of the data, including the extent of missing information. Although we are obviously not in a position to collect further information and do not intend to impute missing values, we believe that balanced discussion of the strengths and limitations of the data will assist in the evaluation of this study.

Risk of baseline imbalance: our analysis concentrates on the two outcomes (exam performance and school attendance) where there are documented outcome data in both years of the study. In our secondary analyses of the impact of health related outcome, we have largely restricted ourselves to areas where there are repeat measurements that allow baseline comparability.

Risk of bias from sequence allocation: as part of our analysis, we will be able to use raw data to appraise the randomisation of schools, comparing their size, location and other baseline characteristics (such as involvement in other ICS programs), which will help to determine if the randomisation process was conducted effectively.

Concerns over use of partial dataset for multiple analyses: due to the study design, various different comparisons between different subgroups were made during the original analysis. Analyses that only make use of subgroups are vulnerable to having inadequate power to detect true effects (type 2 errors – false negative) and performing multiple analyses increases the chance of making type 1 errors (false positive). We aim to deal with this problem by making a small number of analyses using as much of the original data as possible at each stage and concentrating initially on the direct intervention effects on the major study outcomes.

Sample size calculation – detectable effects from this study.

In advance of performing this analysis, we have estimated what size effects we believe can plausibly be detected (power to detect = $1-\beta \geq 80\%$) with this study size, at different significance levels (α). We have based these estimates on the assumption of complete data for attendance outcomes in schools with 400 pupils/school and 25 schools/cluster, and a baseline rate of attendance of 0.72% (ie 28% absent) and a coefficient of similarity within schools (k) of 0.25, which represents a moderate degree of within-cluster similarity. This also includes an estimated adjustment for the stepped-wedge design. The original analysis estimated an approximate 7% overall improvement in attendance (ie from 28% to 21% absent) associated with the intervention.

| Table 2 % improvement in school attendance | Power to detect effect at differing significance level (α) | | | |
|---|---|---------------|---------------|----------------|
| | $\alpha=0.1$ | $\alpha=0.05$ | $\alpha=0.01$ | $\alpha=0.001$ |
| 3% | <70% | <70% | <70% | <70% |
| 5% | ~80% | ~70% | <70% | <70% |
| 7% | >90% | >90% | 80-90% | ~70% |
| 9% | >90% | >90% | >90% | ~80% |
| 11% | >90% | >90% | >90% | >90% |

On this basis, we feel this dataset is likely to be of adequate size to detect an effect size as found by the original study at standard levels of significance, but would be unlikely to detect an effect that was smaller than this. Larger effects should be straightforward to detect.

3. Planned replication work – data preparation and statistical methods

3.1. Pure replication

For the pure replication, we aim to perform the following:

- Re-conduct the original analyses, using the raw data (as provided to the authors by ICS) and do files (as provided by the authors).
- Compare the outputs obtained with those in the published data tables.
- Provide a commentary on the use of statistical functions in the original analysis.
- Explicitly describe how the original study took account of the clustered nature of the data.

3.2. Data preparation for statistical replication

3.2.1. Data preparation for the two major study outcomes

This will be performed as follows:

1. School attendance. The original study describes measurement of school attendance being performed by ICS fieldworkers making unannounced school visits to check the presence of individually named children, with each pupil in grades 1-7 receiving an average of 3.8 visits (p179). From the raw data, it appears that there was variation in the number of visits performed per child / per school with (1998 data only, only pupils with >0 observations) a mean of 3.55, median 4, range 1-5.

We will prepare this data in the form of a binary outcome of whether individual child was present at the ICS various school visits. We will take account of the within child-clustering of results due to repeat observations of the same individuals.

The original study described school attendance being measured between May 1998-March 1999 (period 1) and May 1999-November 1999 (period 2) ((1) p195). Western Kenya's climate is highly seasonal, with wet seasons in approximately March-May and October-November. Additionally, families will often travel around the time of the Christmas vacation in December. On this basis, we feel there is likely to be a seasonal pattern to school attendance, and it would therefore be important to measure attendance in the same calendar months in period 1 and period 2 to avoid confounding by seasonal variations. We therefore aim to determine a proportion of school visits in year 1 (1998) and to compare this to the same parameter in year 2 (1999). The dates for the school visits are recorded with the raw data from ICS – we will use this to determine which visits correspond to which year of the study, and also examine if school visits were evenly distributed between different study arms.

2. Exam performance. The original study describes exams in English, Maths and Science-Agriculture administered by ICS for pupils in Grades 3 to 8 in all schools ((1)). Pupils in grades 1 and 2 did not take these exams, so will be excluded from this analysis. From the raw data, we aim to perform the same initial data transformation steps as described in the original study. This would create a normal distribution of marks across all schools for a particular subject and age-group with a mean of zero and standard deviation of one. For each pupil assessed in each exam, we would then transform their individual mark into a measure of deviation from the exam-specific mean (z-score). We would sum these measures across all exams to produce individual summary z-score statistics. We will only examine combined results for average scores across all three exams.

3.2.2. Data preparation for the intervention (principally drug treatment) will be as follows

Baseline comparability

We will produce a summary table describing the characteristics of the three intervention groups, including descriptions of cluster-level (eg school size) and individual-level (eg age) variables.

Compliance with treatment

In our primary analyses, we will conduct this as an Intention To Treat (ITT) analysis. This means that we will not take account of reports of compliance with drug treatment in the study (reported to be approximately 72% amongst eligible pupils), although presumably pupils were not able to “opt-out” of the educational components of the intervention. This study was clearly intended as an evaluation of the “real-world” effectiveness of this intervention rather than a tightly-controlled efficacy study, so it would not be appropriate to attempt to conduct a “Per Protocol” (PP) analysis of these data where only those individuals that actually received the intervention are evaluated. We accept that use of an ITT analysis format may “dilute” the true effectiveness of the intervention, but this also makes it a fairer evaluation of the effectiveness that the intervention is likely to achieve in actual usage. Furthermore, if there is a substantial within-school indirect benefit (as described by the original study authors), then all pupils in the interventions schools will have detectable benefits.

Attrition (Drop-out) from study

An important potential source of bias in any randomised trial is differential drop-out (attrition) between different arms of the study. We will examine the data to determine the extent and type of differential attrition, and depending on our findings, we may apply an appropriate adjustment to estimates of intervention effect and/or discuss the possible consequences that may have on the reliability of the estimates.

Eligibility for treatment

In the original study design, there was a clear intention that girls over the age of 12 years should not receive the drug component of the intervention, although, in practice, some did actually receive it (approximately 10%). We will consider that only male pupils and girls of 12 years or less were eligible for participation in the drug treatment aspect of the study and we will restrict our estimation of the direct effects to comparison of these groups between schools.

Separate analyses by treatment type

As previously described, we aim to perform separate analyses for the two different types of intervention treatment that were used – albendazole for treatment of geohelminths and praziquantel for treatment of schistosomiasis. We will assume that schools in Group 2 who received schistosomiasis treatment in the 2nd year would have been eligible for this if they had been tested in the first year – these will form the control schools for this comparison. In the absence of data on schistosomiasis prevalence on schools in Group 3, we will not be able to include data from this group in analysing the effects of the schistosomiasis treatment.

3.3. Data handling in main analysis

We will approach the analysis of the data, as much as is possible, by using standard methods used in the analysis of cluster randomized trials in the medical literature. We will follow from Hayes+Moulton’s book *Cluster Randomized Trials* (7). Methods of analysis of cluster randomized trials account potential correlation between individuals sampled from the same cluster, i.e. the likelihood that two children from the same school are more alike than two children from different schools.

In a cluster randomized, stepped wedge, trials, clusters move from control to intervention conditions in a number of steps. This is the case for our data where there are two steps and the schools fall into one of three categories: received the intervention in 1998 and 1999; only received the intervention in 1999; and did not receive the intervention in either 1998 or 1999. The design is shown in the schematic below:

| | 1998 | 1999 |
|-------------------|--------------|--------------|
| Group 1 (n=25) | Intervention | Intervention |
| Group 2 (n=25) | Control | Intervention |
| Group 3 (n=25) | Control | Control |

The data from a stepped wedge trial can be thought of as a one-way cross-over, and treated as such, by comparing before and after in the cross-over schools (group 2) and accounting for the secular trend using the non-crossing schools (groups 1 and 3). However, such an approach requires assumptions about the uniformity of the trend and the ability of the model to capture the secular change, and as such loses the advantage of randomization. Also, since we do not have before-intervention data from group 1, the coefficient of effect would come only from a comparison between group 2 schools in 1998 and the same schools in 1999. An alternative is to focus on within-year comparisons between children in schools that have received the intervention and children in schools that have not. This is sometimes referred to as a ‘vertical approach’ to the analysis as the comparison takes place across the columns in the diagram above. We will employ the vertical approach in the analysis of the data.

For the primary analysis of school attendance we will compare observations of attendance or non-attendance across treatment arms, within years. Each child, in each school, will have a number of observations that are either ‘present’ or ‘absent’ and coded as 1 and 0, respectively. Therefore, this analysis will use logistic regression to model the effect of treatment condition on the outcome at each observation. We will include a ‘treatment’ variable in the model that will take the value ‘1’ if the child under observation was enrolled at a school receiving treatment in that year and ‘0’ if the child was in a school not receiving treatment in that year. The primary result will be an odds ratio that a child is present between treatment and non-treatment arms.

We will account for the correlation between repeat observations of the same child and children within a school by including a random effect for the school. Our analysis will initially look within each year, i.e. for 1998 and then for 1999. We will then combine the estimates of effect from the two years, accounting for the correlation in outcomes between the years due to the fact that the same children are measured in each year.

For the secondary analyses, such as estimating effects of the intervention on educational outcomes, we will use an equivalent method to the one used above but with a regression model that is appropriate for quantitative outcomes, i.e. ordinary regression.

The reporting of our analysis will take four steps:

1. Summarize and display the outcomes clearly for each intervention arm in each year. For example, the proportion of children absent in the 25 schools in each group in 1998, and in 1999.
2. Perform an-individual level analysis of the effect of the intervention status within a given year on the outcomes using regression models with random effects to account for clustering. We will report odds ratios and regression coefficients for intervention effect.
3. Combine the estimates of effect across the two years, accounting for correlation.
4. Report results of any adjustment by covariates that are imbalanced at baseline. We will make adjustment for covariates if preliminary inspection of the data suggests that there is imbalance between the arms. We will include covariates in the regression models and report the adjusted estimates.

3.4. Other general comments for planned analysis

Missing data and transfers between schools

Handling of missing data is a particularly difficult issue in analyses of operational studies performed in developing countries – there are no fixed rules on how best to do this. As a general principle, we aim to be as explicit as possible about the extent and treatment of missing data. We will report the number of missing data points for important variables by intervention arm. While the absolute extent of missing data is a concern for the power and generalizability of the study, we will pay particular attention to the relative magnitudes of missing data between the study arms because of concerns about bias. Where there are imbalances in the extent of missing data we will address this in our discussion of the results. We do not aim to perform any imputation of missing data.

There is also the issue of children transferring between schools for different years of the study – if children transferred into or out of school that differed in treatment allocation, this could influence the result of the study. The extent of transfers appears to be limited: approximately 8% of pupils transferred to a different school in the course of the study (see table IV, p172) – we do not suspect that transfers were either caused or affected by study treatment arms this was we feel that excluding these children from the analysis would be unlikely to influence the overall results of the analysis.

Tests for interaction of effects

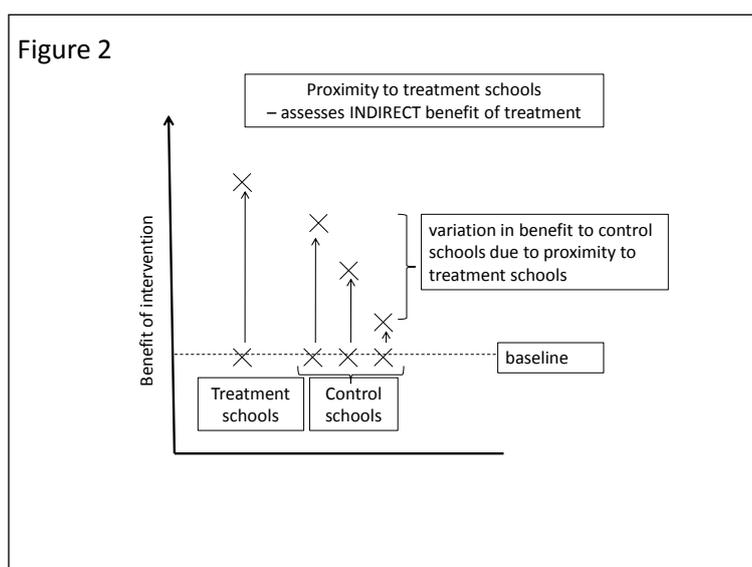
We will test for interactions of any detected effects (either direct or indirect) by both age group and sex only. We will also investigate whether there is any evidence that year of treatment (ie first v second year) had different effects.

Spatial estimation of indirect effect

We will use the same spatial approach employed by the original study for estimating the indirect effect of the intervention. Similar approaches have been used to assess the indirect effects of cholera vaccine in reanalyses of data from an individually-randomised trial in Bangladesh (10, 11) – like worm infections, cholera is transmitted by faecal contamination in the local or household environment. However, this analytic approach is rarely used in cluster-randomised trials as these normally seek to avoid spillover of effects from intervention to control groups, formally known as contamination, which reduces the measured direct benefit of the intervention.

As described earlier, the original study used the observation that schools were located at different physical proximities to treatment schools to estimate the variation between maximum and minimum indirect effect – see figure 2.

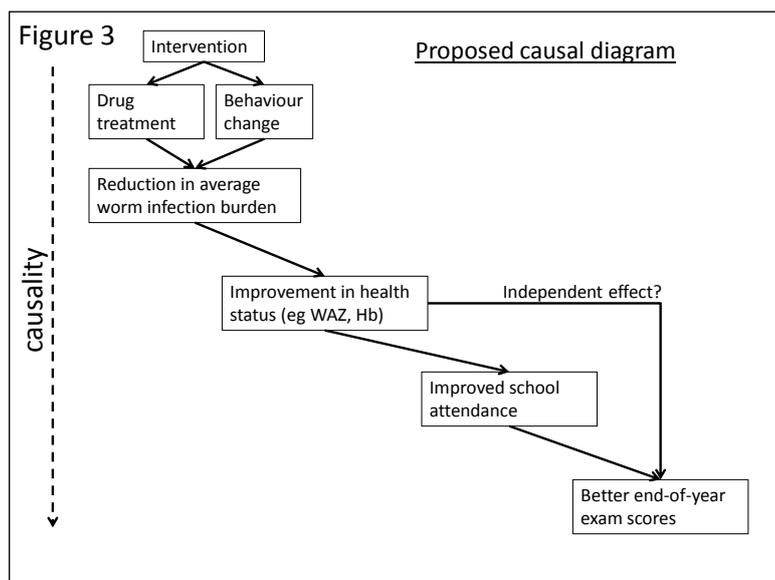
The original study used the number of pupils attending treatment schools within specific distances of the school (p176) and a separate term to describe the total number of primary school pupils within the same zones, to account for factors relating to local population density. We propose to perform the replication using the same analytic approach as described in the original paper, but only if the primary aim demonstrates a direct effect of the intervention.



Use of a causal framework

For both the estimation of the direct and indirect effects of the intervention, we hypothesise a causal pathway for the relationship between the intervention and the eventual outcomes (school attendance and exam performance), which will guide us as to which variables should be used in each level of modelling – see figure 3.

For example, if we were using this causal model to determine whether the intervention had an effect on end-of-year exam performance, we would not include parameters relating to school attendance during the academic year in the model, as this is between the intervention and end-of-year exam performance in the causal model. Similarly, if we wanted to know whether the intervention had an effect on exam performance that is independent of school attendance, we would include both school attendance and exam performance in the model – any residual effect associated with exam scores would then represent an independent pathway. These types of causal models are widely used in epidemiological studies of causation – as discussed in greater detail by Hernán et al (12)



We will seek to determine to what extent the data provides support evidence for a causal relationship between the intervention and school attendance/exam performance, with reference to the Bradford-Hill criteria that are widely used for assessing evidence of causality in the medical literature. We will also discuss what alternative relationships might account for the findings of the original analysis and our own replication. Introduction of these “theories of change” that were not used in the original paper essentially constitutes a “scientific” replication, as per the terminology of the 3ie replication study.

Software use

We will use STATA (v12.0 Statacorp, College Station, Texas, USA) for all statistical analyses. We will provide annotated .do and .log files of all our analysis steps to reviewers and the original study authors.

Report format

We aim to produce a final report with the layout, language and approximate length that is normally used in the medical literature for reporting of clinical trials, including comparison with the results of the original analysis. When reporting confidence intervals, we will use a 95% confidence interval only, and we will report the actual p-values for all statistical tests performed.

Table 2: Biology of worm infections under analysis

| Worm type | Life cycle | Lifespan (all lengths are approximate) | Thresholds for moderate infection (eggs/g faeces) | Treatment used in study |
|---|---|---|--|--|
| Geohelminths (soil-mediated) | | | | |
| Hookworm (<i>N. americanus</i>) | egg in faeces →soil →larvae hatches → skin contact → gut | adult lives 1-10yrs in gut, larvae live in moist soil for 2yrs | 2,000 (WHO) 750 (13)* | albendazole 600mg ('98) / 400mg ('99) given every 6 months |
| Roundworm (<i>A. lumbricoides</i>) | egg in faeces →soil →oral intake of egg →gut | adult lives 1yr in gut eggs last 1-3yrs in soil | 5,000 (WHO)* | |
| Whipworm (<i>T. trichiura</i>) | egg in faeces →soil →oral intake of egg →gut | adult lives up to 5 yrs in human gut | 1,000 (WHO) 400 (13)* | |
| Schistosomiasis (water-mediated) | | | | |
| Schistosomiasis (<i>S. mansoni</i>) | egg in faeces → freshwater → snail → skin contact → veins around gut | adult lives 4+yrs in veins, lifespan in freshwater snail 1yr | 100 (WHO) 250 (13)* | praziquantel if ≥30% prevalence in school |

All information in the above table is drawn from Manson's Tropical Diseases (22nd edition, 2009), unless otherwise referenced.

*= threshold for moderate infection as described in the original study (p167). The citation provided by the authors to justify their use of these thresholds for moderate infection (13) does not appear to provide any clear reason for these choices. We therefore plan to adopt the most recent WHO thresholds for classifying these infections (see p25, Helminth Control in School-age Children, WHO, 2002 (14)).

References

1. Miguel E, Kremer M. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*. 2004;72(1):159-217.
2. Kremer M, Moulin S, Namunyu R. Unbalanced Decentralization: Results of a Randomized School Supplies Provision Program in Kenya. Mimeo Brookings Institution, Washington DC. 2002.
3. CCC. 2012 [cited 2012 July 19]; Available from: <http://www.copenhagenconsensus.com/Projects/CC12/Outcome.aspx>.
4. Taylor-Robinson DC, 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*. 2012;7:CD000371. Epub 2012/07/13.
5. Hawkes N. Deworming debunked. *BMJ*. 2012;346:e8558. Epub 2013/01/04.
6. London Declaration on Neglected Tropical Diseases. Uniting to Combat Neglected Tropical Diseases; London2012.
7. Hayes RJ, Moulton LH. *Cluster Randomised Trials*: Chapman&Hall/CRC; 2009.
8. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276(8):637-9. Epub 1996/08/28.
9. Moulton LH, Golub JE, Durovni B, Cavalcante SC, Pacheco AG, Saraceni V, et al. Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention. *Clin Trials*. 2007;4(2):190-9. Epub 2007/04/26.
10. Root ED, Giebultowicz S, Ali M, Yunus M, Emch M. The role of vaccine coverage within social networks in cholera vaccine efficacy. *PLoS One*. 2011;6(7):e22971. Epub 2011/08/11.
11. Ali M, Emch M, von Seidlein L, Yunus M, Sack DA, Rao M, et al. Herd immunity conferred by killed oral cholera vaccines in Bangladesh: a reanalysis. *Lancet*. 2005;366(9479):44-9. Epub 2005/07/05.
12. Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002;155(2):176-84. Epub 2002/01/16.
13. Brooker S, Miguel EA, Moulin S, Luoba AI, Bundy DA, Kremer M. Epidemiology of single and multiple species of helminth infections among school children in Busia District, Kenya. *East Afr Med J*. 2000;77(3):157-61. Epub 2003/07/16.
14. WHO. *Helminth control in school-age children: A guide for managers of control programmes*2002.