International Journal of Epidemiology, 2015, 1581–1592
doi: 10.1093/ije/dyv128
Advance Access Publication Date: 22 July 2015
Original article

Deworming Programmes, Health and Educational Impacts

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

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

[1]Department of Social and Environmental Health Research, and [2]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

*Corresponding author. Centre for Evaluation, Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK. E-mail: alexander.aiken@lshtm.ac.uk

Accepted 30 March 2015

Abstract

Introduction: Helminth (worm) infections cause morbidity among poor communities worldwide. An influential study conducted in Kenya in 1998–99 reported that a school-based drug-and-educational intervention had benefits for worm infections and school attendance.

Methods: In this statistical replication, we re-analysed data from this cluster quasi-randomized stepped-wedge trial, specifying two co-primary outcomes: school attendance and examination performance. We estimated intention-to-treat effects using year-stratified cluster-summary analysis and observation-level random-effects regression, and combined both years with a random-effects model accounting for year. The participants were not blinded to allocation status, and other interventions were concurrently conducted in a subset of schools. A protocol guiding outcome data collection was not available.

Results: Quasi-randomization resulted in three similar groups of 25 schools. There was a substantial amount of missing data. In year-stratified cluster-summary analysis, there was no clear evidence for improvement in either school attendance or examination performance. In year-stratified regression models, there was some evidence of improvement in school attendance [adjusted odds ratios (aOR): year 1: 1.48, 95% confidence interval (CI) 0.88–2.52, $P=0.147$; year 2: 1.23, 95% CI 1.01–1.51, $P=0.044$], but not examination performance (adjusted differences: year 1: −0.135, 95% CI −0.323–0.054, $P=0.161$; year 2: −0.017, 95% CI −0.201–0.166, $P=0.854$). When both years were combined, there was strong evidence of an effect on attendance (aOR 1.82, 95% CI 1.74–1.91, $P<0.001$), but not examination performance (adjusted difference −0.121, 95% CI −0.293–0.052, $P=0.169$).

Conclusions: The evidence supporting an improvement in school attendance differed by analysis method. This, and various other important limitations of the data, caution against

over-interpretation of the results. We find that the study provides some evidence, but with high risk of bias, that a school-based drug-treatment and health-education intervention improved school attendance and no evidence of effect on examination performance.

Key words: : Helminth, worms parasitic, randomized control trial, primary schools, Kenya

Key Messages

- It remains controversial whether or not deworming school children results in better school attendance—one study conducted in Kenya in 1998–99 remains central to the debate.
- The original study used a complex intervention combining health education and deworming drugs deployed in a cluster quasi-randomized stepped-wedge trial with direct observation of school attendance.
- In this statistical replication, using modern epidemiological methods to examine the same dataset, we found substantial amounts of missing data and that the effects on school attendance varied according to format of analysis, with concerns about the validity of estimates combining data from 1998 and 1999.
- Evidence of health-related secondary outcomes linking the removal of worm infections to possible improvements in school attendance was lacking, making alternative, behavioural, pathways also plausible.
- This re-analysis finds that the original data provide some evidence, with a high risk of bias, that a school-based drug-treatment and health-education intervention improved school attendance, and no evidence of effect on examination performance.

Introduction

Helminth infections cause substantial morbidity across much of the developing world[1,2] and are simple to treat with low-cost medications.[3] Opinions differ over whether treating helminth infections also improves school attendance and educational achievement. One cluster quasi-randomized trial conducted in Kenya in 1998–99[4] is central to the debate.

This report forms the second stage of a re-analysis (or replication) of this influential study. We report the results of a 'pure replication' of the original analysis in a companion paper[5] where we reproduce the original methods used. This paper reports the results of a 'statistical replication' using the same original data. Here we use alternative, pre-specified, methods for data handling and analysis, in line with modern epidemiological approaches (CONSORT statement for cluster-randomized trials[6,7]). We focus on the 'naïve' results of the original study (as described in the pure replication), specifying school attendance and examination performance as the co-primary outcomes, as these were the major focus of the original study.

Methods

Trial design

The cluster quasi-randomized stepped-wedge trial was conducted in primary schools in two districts in Western Kenya between January 1998 and December 1999. In January 1998, there were 92 primary schools in these districts, of which 75 were included in the study (see Figure 1). Schools were systematically allocated, or 'quasi-randomized' (for details see companion paper[5]), into three groups with 25 schools per group by Edward Miguel, Michael Kremer and Sylvie Moulin on an Excel spreadsheet (unpublished observation Edward Miguel). The intervention was introduced over 2 years: Group 1 schools received the intervention in both years; Group 2 schools received the intervention in year 2 only and were in the control arm in year 1; and Group 3 schools did not receive the intervention in either year (Figure 2). The participants were not blinded to their allocation nor administered placebo. Concurrently, Internationaal Christelijk Steunfonds (ICS) were also evaluating five other interventions under their 'School Assistance Programme' in 27/75 study schools (SAP schools).

The intervention

This complex school-based health-education and drug-treatment intervention was delivered by ICS. The health education consisted of regular public lectures and wall charts, and one teacher per school was trained by staff at the Kenya Ministry of Health Division of Vector Borne Diseases (DVBD) to deliver health messages. Trial staff gave 10–15-minute presentations on worm infection

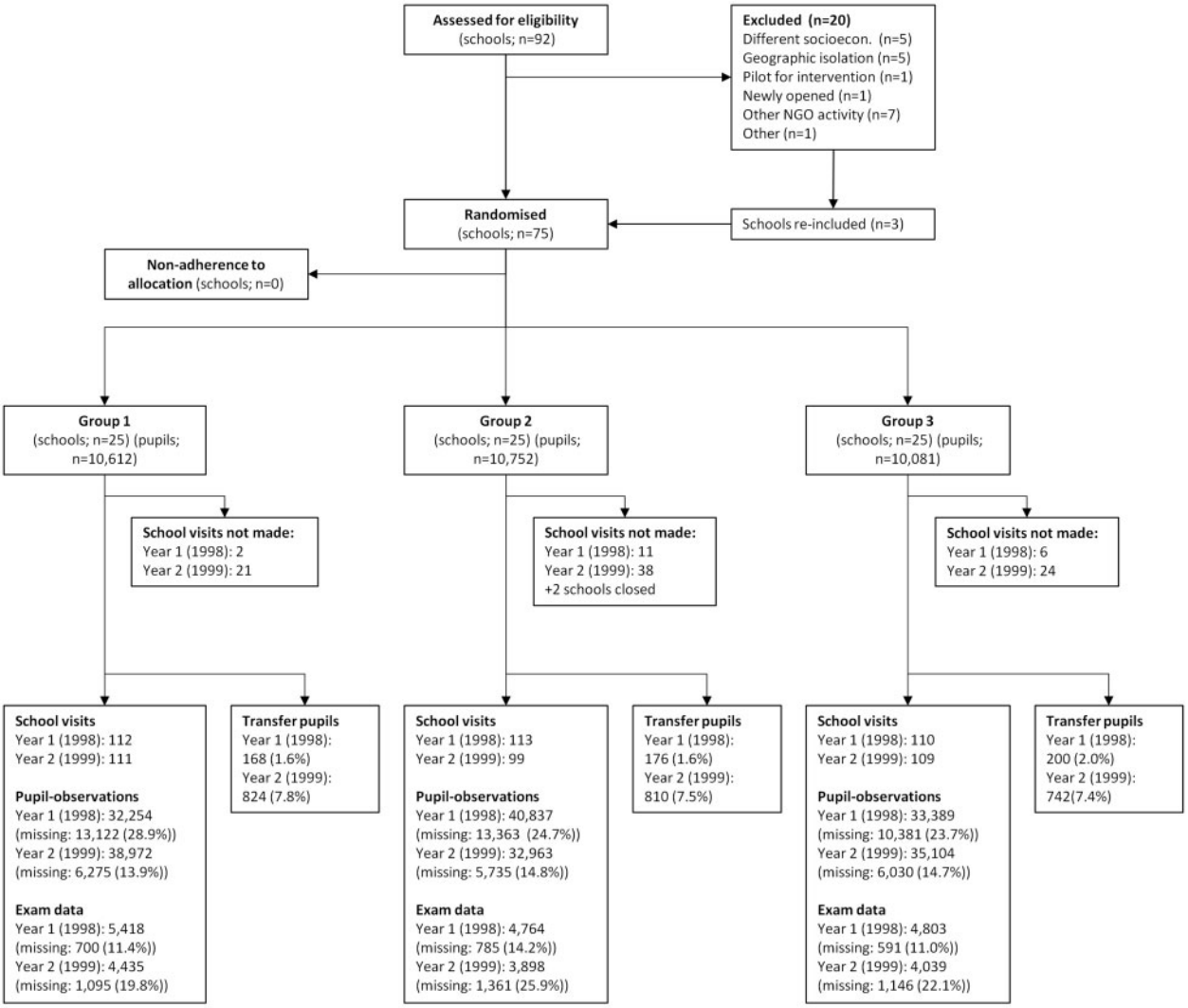

Figure 1. CONSORT diagram.

Study participants were enumerated at the start of the study if they were registered at school in grades 1–8 at the start of 1998. Follow-up is shown for primary outcome data. 'Pupil-observations' refers to the number of times that any pupils were observed in the group, excluding observations after transferring schools. The levels of missingness for pupil-observations is calculated for visits that took place, and therefore does not incorporate missingness due to schools not being visited. Missingness for examination data is based on the number of pupils in standards 3–8 who had not moved school. Socioecon., socioeconomic status; NGO, non-governmental organization.

prevention several times a year. Drug treatment entailed biannual albendazole treatment in schools with geohelminth infection over 50%, and annual mass treatment with praziquantel in schools with schistosomiasis prevalence over 30%. Drug treatment was delivered to all boys and girls aged ≤ 12 years in eligible schools. The DVBD conducted parasitological surveys in advance of delivering the drug treatments that were used to measure the secondary outcomes of the trial. Schools in the control arm received neither component of the intervention. Mass treatment and whole-school health education necessitated school-level randomization. We inferred, in the absence of a protocol, that the complex intervention was intended to be delivered from the start of each calendar/academic year.

Primary outcomes

We specified two co-primary outcomes: school attendance and educational attainment in end-of-year examinations. Outcomes were assessed among a closed cohort consisting of all pupils who were registered in grades 1–8 at the start of year 1 and were eligible for drug treatment (boys and girls aged ≤ 12 years). Outcome data were censored after pupils moved schools. School intervention status was not concealed from fieldworkers collecting the outcome data.

School attendance was measured by ICS fieldworkers during unannounced school visits. Each year was divided into eight potential visit-periods. The 27 SAP schools were scheduled for visits during six periods in years 1 and 2; the 48 non-SAP schools were scheduled for four visits in

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 2. Stepped-wedge design.

Stepped-wedge design shown in schematic form. The intervention was rolled out in 'steps', with Group 1 receiving the intervention in year 1, Group 2 in year 2 and Group 3 in the year after the study.

year 1 and five in year 2. The school visit schedule was concealed and different in each period, although it is unclear how specific dates for visits were chosen. Attendance was binary: pupils were 'in attendance' if observed to be present, and were 'not in attendance' if they were not present. If the observation data were missing but records indicated the student had 'dropped out', then a pupil was coded as 'not in attendance'. Handling of missingness in the attendance measure is described in Appendix 1 (available as Supplementary data at *IJE* online). In our individual-level analysis, each attendance observation was treated as a binary outcome and we did not aggregate observations for each pupil. Examinations in Mathematics, English, and Sciences were administered by ICS for pupils in grades 3–8 at the end of each year. The raw marks were transformed into a measure of deviation from the examination-specific mean (z-score), and averaged for each pupil.

Secondary outcomes

Worm infection and weight-for-age (WAZ) and height-for-age (HAZ) data were collected and specified in this analysis as secondary outcomes. These data were available for comparison between arms in year 1 only.

At baseline and the start of year 2, the DVBD assessed worm infection prevalence among sub-samples of pupils from schools about to initiate treatment. No testing was performed for Group 2 at baseline or Group 3 at any point. It is unclear how these sub-samples of pupils were selected. In Group 1, a 'representative subset' of the pupils tested at baseline was sought for re-testing in year 2.[4] The Kato-Katz technique was used for sample preparation, and egg counts from two readers were averaged and converted into eggs per gram of stool values. Arithmetic mean egg counts with standard World Health Organization (WHO) thresholds for moderate infection were calculated.[8]

ICS collected anthropometric measures from all pupils in grades 3–8 at baseline and the start of year 2. A questionnaire was administered on a pre-announced day and only to pupils who were present. A single enumerator read the scales and took height measurements for all of the pupils at a visit. Pupils were asked their age and ICS staff were encouraged to cross-check against school records. WAZ and HAZ were converted to z-scores by the original authors. WAZ and HAZ data were only considered to be missing if they were not recorded for grade 3–8 pupils.

Ethics and consent

In both years, community and parent meetings were held in intervention schools immediately before delivery of the intervention. In year 1, parents who did not wish their children to receive the drug treatment were asked to inform their school headmaster. In year 2, under recommendation from the Kenyan Ministry of Health, ICS was required to collect written consent from parents for children to receive drug treatment. Pupils in all arms were asked for their consent to take part in the questionnaire survey. It is unclear what informed consent procedures were carried out for attendance observations in schools.

Statistical analysis

Analyses included those eligible for treatment with deworming drugs: all boys and girls aged ≤ 12 years. All analyses were performed according to the original assigned group (intention-to-treat). In accordance with our interpretation of the intention-to-treat of this complex intervention, school attendance observations of pupils in 1998 were assigned to the treatment condition in Group1 and the control condition in Groups 2 and 3, and in 1999 observations were assigned to the treatment condition in Groups 1 and 2 and control condition in Group 3.

First, we described the characteristics of the study population and investigated patterns of missing data. We calculated mean cluster-summaries of baseline characteristics with confidence intervals for each group. We calculated mean age for each grade and applied this to pupils with missing age data as a simple form of imputation. We calculated the between-cluster coefficient of variation (k) for school attendance in the 50 control schools in year 1 and in the 50 intervention schools in year 2, and similarly the intra-cluster coefficient of variation (ICC) for examination performances. Initial analyses identified an unexpected cluster-level association between the level of school attendance and the total number of pupil-observations performed, which was influenced by whether or not schools were involved in the SAP programme. To describe and investigate this association further, we plotted the proportion of pupils observed as present in each school against the number of observations made in a school, stratified by year

and by allocation group, and fitted ordinary-least-squares regression lines.

The primary-outcome analyses were conducted in three steps that increased progressively in complexity to reflect the cluster-allocated stepped-wedge design of the trial. First, in each year the means of the school-level summary outcomes for each group were calculated, and also for each intervention arm. The latter were compared within years using the unpaired t-test. Second, year-stratified random-effects logistic regression models were used to examine the association between the intervention and attendance, and mixed-effects linear regression for examination performance (see Appendix 2 for model specification, available as Supplementary data at *IJE* online). Random effects were fitted for school. Third, regression models with a fixed effect for year were used to combine the 2 years, which implicitly included comparison between Group 2 in year 1 (control) in year 2 (intervention). *P*-values for logistic regression were calculated using likelihood ratio tests. Sensitivity of the results to the exclusion of one school with no pupils recorded present in year 2 was examined.

All regression models included terms for the population size of the school and the zone of the school, since these were used to stratify the quasi-randomization, and further adjustment was made for variables that showed imbalance between groups at baseline. Using the combined-year logistic regression model, we investigated potential interaction by age and by school SAP status; the latter was not prespecified in the pre-analysis plan. For the secondary analysis, we compared mean cluster summaries for each arm using unpaired t-tests.

In analyses that were not pre-planned, we investigated the sensitivity of our school attendance results to the assumption of the intention-to-treat applying from the start of each year. This was based on information in the study timeline (Appendix 4, available as Supplementary data at *IJE* online) that indicated that the drug component of the intervention was not delivered at the start of each year. We investigated two scenarios. In scenario one, we excluded observations of attendance in the first visit-period in 1998, and added observations in the first two visit-periods in 1999 to the analysis for the first year, assigning observations in Group 2 during both these visit-periods to the control condition. Therefore, 'year 1' comprised observations in the second to the eighth visit-periods in 1998 plus observations in the first and second visit-periods in 1999. 'Year 2' comprised observations in the third to the eighth visit-periods in 1999. This approximates to what was done in the original analysis.

In scenario two, we excluded observations of attendance in the first visit-period, and also excluded the observations in the first two visit-periods in 1999. This data handling avoids comparing observations of pupils in different years in the year-specific analyses, and is our preferred method to accommodate the described timing of the drug treatment.

Results

The trial took place between January 1998 and December 1999 (map Appendix 3; timeline Appendix 4; available as Supplementary data at *IJE* online). All 75 schools agreed to take part and none dropped out. Approximately 10500 pupils were enrolled in each group (Table 1). A substantial proportion of age data were missing at baseline (6646 pupils, 21.1%); after imputation (using the mean age in each grade), there were 702 pupils with missing age data (2.2%)—these pupils were also missing grade data. Pupils in Group 1 were 0.4 years older than those in Group 2 and a higher number of Group 2 schools were enrolled in the SAP ($n = 12$) than in Group 1 ($n = 7$) or Group 3 ($n = 8$). The sex ratio, proportion eligible for drug treatment, mean WAZ and distance to Lake Victoria were balanced across the groups. There were substantial missing sex data (3399 missing sex observations/31 445 pupils; 10.8%) and the amount of missing data varied between groups. The mean school size was similar in the three groups but the range was much larger for Group 2 (minimum 37; maximum 1392).

Pupils moved schools during the trial with approximately the same frequency in each group. During the first year, 168 pupils (1.6%) in Group 1 moved to a different school, 176 pupils (1.6%) in Group 2 and 200 (2.0%) in Group 3. By the end of year 2, 824 pupils (7.8%) in Group 1 moved, 810 (7.5%) in Group 2 and 742 (7.4%) in Group 3.

In year 1, two Group 2 schools were temporarily closed and pupils from these schools were absorbed by other local schools. Both were reported to have re-opened, but for one school none of the 7 pupil-observations in year 2 were recorded as present. This school was included in later analyses, but a sensitivity analysis showed no major impact of excluding it. In addition, one school in Group 3 had no attendance observations recorded in year 1 and one school in Group 2 had no examination results in year 2.

All intervention schools were eligible for mass albendazole treatment. In year 1, 6616 pupils in Group 1 received drug treatment (72.1% of those eligible), with none receiving treatment in Group 2 or 3. In year 2, 4516 Group 1 pupils (52.1%) and 4159 Group 2 (47.5%) pupils received drug treatment, as well as 91 pupils in Group 3. A minority of schools were eligible for mass schistosomiasis treatment (6/25 in year 1, 16/50 in year 2).

Table 1. Baseline characteristics

		Group 1	Group 2	Group 3
Pupil-level characteristics				
Number of pupils		10612	10752	10081
Male, or female and ≤ 12: year 1 (1998)		9180	9299	8660
Male, or female and ≤ 12: year 2 (1999)		8661	8749	8172
Missing eligibility data (%)		187 (1.8)	303 (2.8)	250 (2.5)
Proportion male (95% CI)		0.53 (0.52 – 0.54)	0.51 (0.47 – 0.55)	0.52 (0.50 – 0.53)
Missing sex (%)		562 (5.3%)	1053 (9.8%)	1784 (17.7%)
Mean age (95% CI)		11.8 (11.6 – 12.0)	11.4 (11.2 – 11.7)	12.3(12.1 – 12.6)
After imputation (95% CI)		11.4 (11.2 – 11.6)	11.0 (10.8 – 11.2)	11.2 (11.0 – 11.4)
Missing age (%)		1662 (15.7)	1929 (17.9)	3055 (30.3)
After imputation (%)		155 (1.5)	334 (3.1)	213 (2.1)
Mean WAZ (95% CI)		−1.38 (−1.44 – −1.33)	−1.45 (−1.53 – −1.36)	−1.44 (−1.52 – −1.36)
Missing WAZ *n/N* (3–8th standard 1998) (%)		1792/6233 (28.8)	1740/5672 (30.7)	1382/5498 (25.1)
School-level characteristics				
Number of schools		25	25	25
School Assistance Programme (SAP) schools		7	12	8
SAP A interventions	Received textbooks in 1996	2	4	1
	Received grants in 1997	2	3	2
	Received grants in 1998	2	2	2
SAP B interventions	Early childhood development	2	7	5
	Teacher incentives	5	5	3
Latrines per 1000 pupils (95% CI)		7.4 (6.1 – 8.8)	6.2 (4.7 – 7.7)	6.6 (5.2 – 7.9)
Km to lake Lake Victoria (95% CI)		10.0 (7.9 – 12.2)	9.9 (6.7 – 13.2)	9.5 (6.9 – 12.0)
Pupils per school [mean (min-max)]		424 (168 – 772)	430 (37 – 1392)	403 (103 – 752)

All pupils who were enrolled or registered in school at the start of year 1 (1998) are included in the denominator. Point estimates and confidence intervals for pupil characteristics were calculated using the means of cluster-mean summary measures. Each SAP school received zero or one of the A interventions and one of the B interventions.

Field documents from year 1 indicate that the educational components of the intervention were delivered to Group 1 schools. Documentation from the second year of the study was not received.

The numbers of school visits conducted are shown in Figure 1. In year 1, 19 planned school visits were not conducted (5.4% of intended total), with the majority (11) in Group 2. In year 2, 83 planned visits were not conducted (20.6% of intended total), again the highest number being in Group 2 (21 in Group 1, 38 in Group 2 and 24 in Group 3). Data were available for 74% of pupils during conducted visits in year 1 and 86% in year 2, and within years the proportions were broadly similar between groups. In all three groups, school attendance was higher in year 1 than in year 2 (Figure 3). In year 1, but not year 2, there were several schools that had more than 95% attendance. All of the schools with attendance above 95% were non-SAP schools. There was an unexpected association between the number of school-attendance observations in a school and the school's mean attendance, which depended on the intervention arm. As indicated by the slope of the lines in Figure 3, in 2/3 intervention group-years school attendance was higher in schools where more observations were undertaken. Conversely, the opposite relationship was seen in all three of the control group-years.

The means of the cluster summaries of school attendance for intervention schools in year 1 and year 2 were both higher than the corresponding control school means, but there was no statistical evidence for the differences (year 1 difference +5.48%, 95% CI -1.48–12.44, t-test $P = 0.12$; year 2 difference +2.16%, 95% CI -3.95–8.27, t-test $P = 0.48$) (see Table 2). These cluster-level risk differences were equivalent to odds ratios (OR) of 1.78 (year 1) and 1.21 (year 2). The coefficient of variation for school attendance was in line with the sample size calculation at 0.17 in year 1 and 0.11 in year 2.

The year-specific logistic regression models indicated weak evidence of an association between the intervention and attendance in year 1 (OR 1.77, 95% CI 0.91–3.44, $P = 0.097$). In year 2 the effect size was smaller but the confidence intervals much narrower, with stronger evidence for an effect (OR 1.23, 95% CI 1.01–1.51, $P = 0.047$). Results were similar after adjusting for SAP status and age.

In the regression analysis with both years combined, we found that the effect size was greater than in either year

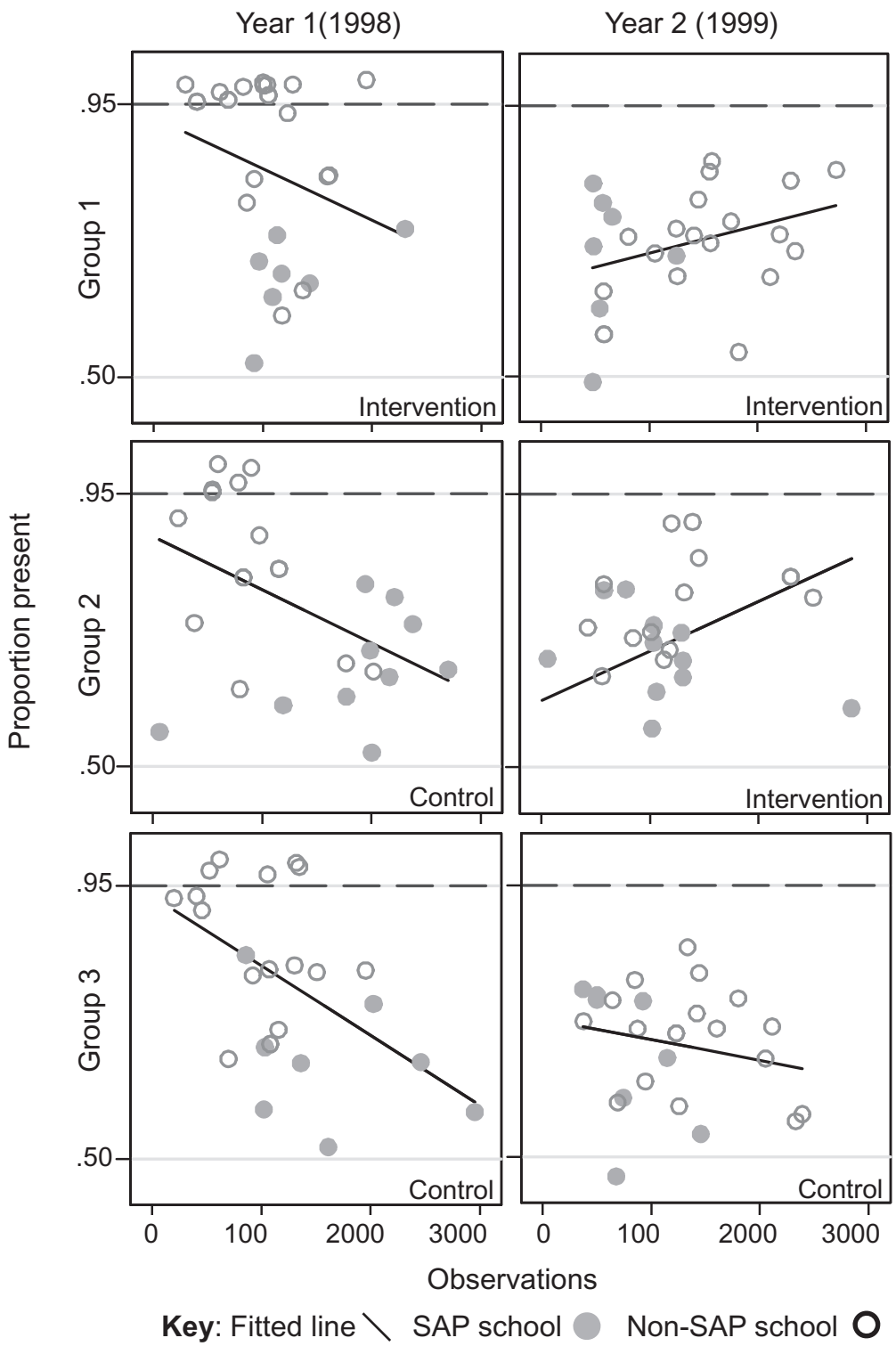

Figure 3. Scatter-plot of proportion present against number of observations, by year and group.

Scatter plots of proportion present against the number of observations in each school by year and by allocation group. The dotted line indicates 95% attendance. Two schools in Group 2 have been excluded from the charts to preserve the scale: one in year 2 (1999) where no pupils are recorded present, and one in year 1 (1998) with a disproportionately large number of observations (approximately 6000).

Table 2. Primary outcomes—results

Cluster summaries								Individual-level random effects (all adjusted for pupil population size and zone)			
Year(s)	Attendance (%)							Unadjusted (*N* 1998 = 90571; *N* 1999 = 89170)		Adjusted for age, SAP (*N* 1998 = 89540; *N* 1999 = 88735)	
	Grp1 % (*N*)	Grp2 % (*N*)	Grp3 % (*N*)	Intervention % (*N*)	Control % (*N*)	Difference (95% CI)	*P*-value	Odds ratio	*P*-value†	adjusted Odds ratio	*P*-value†
1998	84.2 (25)	77.0 (24)	80.4 (24)	84.2 (25)	78.7 (48)	5.48* (−1.48 – 12.44)	0.121	1.77 (0.91 – 3.44)	0.097	1.48 (0.88 – 2.52)	0.147
1999	72.5 (25)	70.9 (25)	69.5 (25)	71.7 (50)	69.5 (25)	2.16** (−3.95 – 8.27)	0.483	1.23 (1.01 – 1.51)	0.047	1.23 (1.01 – 1.51)	0.044
1998 + 1999								1.78 (1.70 – 1.87)	<0.001	1.82 (1.74 – 1.91)	<0.001
								Age-stratified results Interaction test *P*-value < 0.001		≤7yrs: 2.43 (2.27 – 2.61) 8–9yrs:1.87 (1.76 – 2.00) 10–11yrs:1.93 (1.82 – 2.05) 12–13yrs:1.53 (1.44 – 1.64) ≥14yrs: 1.42 (1.32 – 1.52)	
								SAP-status stratified results Interaction test P-value = 0.045		SAP: 1.88 (1.78 – 2.00) Non-SAP: 1.74 (1.63 – 1.86)	
	Mean ICS examination score							Unadjusted (*N* 1998 = 12011; *N* 1999 = 9830)		Adjusted for age, SAP (*N* 1998 = 11999; *N* 1999 = 9826)	
	Grp1 Mean (*N*)	Grp2 Mean (*N*)	Grp3 Mean (*N*)	Intervention Mean (*N*)	Control Mean (*N*)	Difference (95% CI)	*P*-value	Treatment coefficient	*P*-value	Treatment coefficient	*P*-value
1998	−0.031 (25)	0.037 (25)	0.118 (25)	−0.031 (25)	0.077 (50)	−0.109 (−0.332 – 0.115)	0.336	−0.131 (−0.321 – 0.058)	0.173	−0.135 (−0.323 – 0.054)	0.161
1999	−0.033 (25)	0.082 (24)	0.052 (25)	0.023 (49)	0.052 (25)	−0.028 (−0.228 – 0.171)	0.777	−0.015 (−0.199 – 0.168)	0.870	−0.017 (−0.201 – 0.166)	0.854
1998 + 1999								−0.117 (−0.292 – 0.058)	0.191	−0.121 (−0.293 – 0.052)	0.169

Cluster summaries are the unweighted mean of school-level summaries. The individual analyses are adjusted for the variables used in stratifying the randomization. The total units of analysis for the combined year 1 and year 2 analyses are the sums of the yearly totals. The random-effect is fitted for the school. Transfer pupils were excluded after moving schools.

*This risk difference corresponds to an odds ratio of 1.78.

**This risk difference corresponds to an odds ratio of 1.21.

† Likelihood ratio test.

considered individually. There was strong evidence of an effect (OR 1.78, 95% CI 1.70–1.87, $P < 0.001$), including after adjustment (aOR 1.82, 95% CI 1.74–1.91, $P < 0.001$). In this combined-year adjusted model, there was strong evidence of an interaction between intervention status and age ($P < 0.001$), with a stronger effect for younger age groups. The intervention effect in SAP schools was aOR 1.88 (95% CI 1.78–2.00) and in non-SAP schools was aOR 1.74 (95% CI 1.63–1.86) with some evidence for an interaction ($P = 0.045$). All results were similar when the school that had no pupils recorded present in year 2 was excluded from analyses.

The results of a sensitivity analysis exploring effects of the handling of the treatment condition on school attendance results are shown in Table 3 (full results in Appendix 5, available as Supplementary data at *IJE* online). In scenario one, 11588 attendance observations performed at the start of 1998 were excluded, and 31404 observations occurring during the first two visit-periods in year 1999 were handled as 'year 1' observations. In this scenario, the cluster summary mean differences were slightly larger in both 'years', and had smaller *P*-values than in our pre-specified analysis. In adjusted regression models, the OR for 'year 1' was slightly closer to the null, whereas the result for 'year 2' was virtually unchanged. The adjusted combined-year logistic regression OR was larger, with similarly strong evidence. In scenario two, 11588 observations at the start of 1998 were excluded, as well as the 31404 observations during the first two visits in 1999. In comparison with our pre-specified primary analysis, the year-specific results were largely unchanged, with the cluster summary mean difference in year 2 being slightly larger. For the combined-year logistic regression analysis, the adjusted OR was larger than in the pre-specified analysis.

Examination data were available for approximately 5000 pupils per group in year 1 and 4000 pupils per group in year 2, with balance in the extent of missing data. The values of ICC for examination performance were large: 0.20 in year 1 and 0.16 in year 2. There was no evidence of an association between intervention and examination performance in the cluster-mean analysis or the individual-level linear regression models (Table 2). We found evidence for a reduction in roundworm and whipworm infection, and a large imprecise difference in schistosomiasis infection between Group 1 and Group 2 at the start of year 2 (Table 4). We found no evidence for effects on either WAZ or HAZ. There was a high degree of missingness in the anthropometric measures.

Discussion

Our re-analysis of data from a cluster quasi-randomized stepped-wedge trial of a complex intervention found some evidence for an improvement in school attendance, but with high risk of bias. This effect differed by age and by whether or not schools were involved in another intervention programme (the School Assistance Programme). The strength of evidence supporting the improvement in school attendance was dependent on the analysis approach used, in particular when the two years were combined. Our first year-stratified analysis using unweighted cluster summaries found higher attendance in intervention schools in both years, but the evidence was weak. This analysis should be robust but may not be statistically optimal. To improve precision, we used random-effects regression on school attendance observations, an approach which gives greater weight to clusters with higher numbers of observations. This analysis found limited evidence of a moderate effect in year 1, and evidence of a small effect in year 2; these effects were of equivalent magnitude to the cluster-summary results. Finally, we pooled data from the 2 years using logistic regression and found strong evidence of an effect that was substantially larger in magnitude than either of the two year-specific effects. There was no

Table 3. Sensitivity analysis: key outcomes for school attendance in primary analysis and alternative scenarios

Result	Primary analysis (95% CI)	Alternative Scenario One (95% CI)	Alternative Scenario Two (95% CI)
Adjustment applied	-	Drop observations at start of year 1, re-code observations at start of 1999 as end of year 1	Drop observations at start of year 1 and at start of year 2
% difference in cluster summaries 1998	5.48 (−1.48 – 12.44)	7.38 (−0.19 – 14.95)	5.91 (−1.35 – 13.17)
% difference in cluster summaries 1999	2.16 (−3.95 – 8.27)	3.57 (−1.33 – 8.47)	3.57 (−1.33 – 8.47)
Adjusted OR 1998	1.48 (0.88 – 2.52)	1.44 (1.03 – 2.00)	1.49 (0.88 – 2.54)
Adjusted OR 1999	1.23 (1.01 – 1.51)	1.22 (0.97 – 1.52)	1.22 (0.97 – 1.52)
Adjusted OR 1998 + 1999	1.82 (1.74 – 1.91)	1.92 (1.82 – 2.01)	2.13 (2.02 – 2.25)

Table 4. Secondary outcomes - worm infections and WAZ and HAZ at start of year 2 (1999)

Worm infection Type of worm infection	Group 2 pupils pre-intervention	Group 1 pupils after 1 year of intervention	Difference (95%CI)	*P*-value
Pupils tested (*n*)	1233	746		
Average egg count (eggs/g; arithmetic mean)				
Hookworm	694	151	−543(−744 – −342)	<0.001
Roundworm	4283	1289	−2994 (−4540 – −1448)	<0.001
Whipworm	374	254	−120 (−386 – 146)	0.367
Schistosomiasis	245	115	−130 (−316 – 56)	0.165
Proportion with moderate infection (WHO thresholds)				
Hookworm	7.8%	1.8%	−6.0% (−8.7 – −3.2)	<0.001
Roundworm	23.6%	7.8%	−15.7% (−23.8 – −7.7)	<0.001
Whipworm	7.6%	6.6%	−1.0% (−7.3 – 5.2)	0.747
Schistosomiasis	17.1%	8.0%	−9.1% (−20.2 – 2.0)	0.107

WAZ and HAZ Group(s)	Number tested	Intervention status	Mean z-score	Difference from Group 1 (95%CI)	*P*-value
Weight-for-age z-score (WAZ)					
1	2981	Received	−1.329	ref	–
2	2097	None	−1.282	0.047 (−0.094 – 0.188)	0.507
3	2195	None	−1.380	−0.051 (−0.190 – 0.088)	0.469
2+3	4292	None	−1.332	−0.003 (−0.119 – 0.125)	0.962
Height-for-age z-score (HAZ)					
1	2982	Received	−1.231	ref	–
2	2098	None	−1.129	0.102 (−0.133 – 0.337)	0.390
3	2196	None	−1.453	−0.222 (−0.455 – −0.010)	0.061
2+3	4294	None	−1.294	0.064 (−0.149 – 0.276)	0.552

The clustered nature of the data was accounted for by calculating the 95% confidence intervals around the mean of the cluster means. There were missing data for individual Group 1 pupils tested for hookworm ($n=2$) and whipworm ($n=4$) and in Group 2 for hookworm ($n=1$), whipworm ($n=6$) and schistosomiasis ($n=3$). WAZ data were unavailable for 2128 (41.7 %) Group 1 pupils in grades 3–8, 2722 (56.5 %) Group 2 pupils and 2448 (52.7 %) Group 3 pupils. HAZ data unavailable for 2127 (41.6 %) Group 1 pupils grades 3–8, 2721 (56.5 %) Group 2 pupils and 2447 (52.7 %) Group 3 pupils.

evidence of effect of the intervention on examination performance.

The trial had several strengths; in particular, this was a large study and the attendance data were collected using direct observation. A limitation of the trial was that it lacked a clearly pre-documented plan for sampling, data collection, data management and analysis. These documents would help to understand why some schools were not visited, the day of the week on which schools were visited and any ways in which the data collected differed from what was intended. Without access to these documents, we cannot explain why the observed relationship between the number of school-attendance observations in a school and the school's mean attendance depended on the intervention arm (Figure 3). We did not anticipate this correlation when pre-specifying our analysis plan. This underlying correlation in the data could potentially bias effect estimates using random-effects regression methods, since these gave more weight to schools with more observations.

The stepped-wedge design appeared to exacerbate the influence of the unexpected patterns in the data. The combined-years model estimated an effect that was higher than either of the two year-specific effects. We suggest this may partly be due to the fact that the intervention effect from the combined-year analysis includes a non-randomized comparison of Group 2 between years, before and after introduction of the intervention. This comparison can usefully increase the precision of the effect estimate if secular trends are adequately controlled.[9] We suggest that a simple diagnostic that should lead to caution in the analysis of stepped-wedge trials because of inadequate control for secular trends might apply when—as we have found in this study—the combined-step effect estimate is substantially outside the bounds of the randomized step-specific estimates; in this study, years represent the steps in the wedge. We are particularly concerned about the reliability of this before-after comparison because, as Figure 3 shows, in Group 2 the association between the number of pupil observations and mean school attendance changed between

years. This would potentially lead to overestimation of the effect on attendance in a weighted analysis. Furthermore, as this is a closed cohort, the study population in year 2 was on average 1 year older than in year 1, some pupils had dropped out and some had aged out (i.e. left school after completing grade 8). Due to limitations in the data collection, we did not attempt to censor pupils who left school. Thus, a pupil observed to be 'absent' on a particular visit was progressively more likely to have permanently left school.

Further concerns arise from the patterns in the data. The schools with very high (> 95%) attendance in year 1 were all non-SAP schools, as shown in Figure 3, and fieldworker visit schedules were different for SAP and non-SAP schools (different frequency and timing; Appendix 2, available as Supplementary data at *IJE* online). In year 2, the attendance patterns of SAP and non-SAP schools were similar, and the visit schedules were more alike. It is possible that these systematic differences in data collection affected the measured level of attendance because of seasonal variation in school attendance. With regard to planned school visits that were not performed by fieldworkers, it is possible that the obstacles to visiting would also have affected pupils and may have varied over time. These effects could lead to bias, especially when combining results across the two study years.

In light of these issues, we are particularly uncertain about the validity of estimates arising from combined-year analyses of these data, and advise that such results should be interpreted with caution. We have greater confidence that the year-stratified analyses reflect attendance differences between treatment and control, but with modest statistical evidence. A limitation of this research is that, to our knowledge, there is no accepted method for combining the randomized comparisons in each year to estimate an overall effect and confidence interval with binary data from a stepped-wedge trial without evoking a non-randomized before-after comparison in clusters that change treatment arm.[7]

In an exploratory analysis, we investigated the sensitivity of the results to the specification of 'year 1' and 'year 2'. In neither of the two scenarios were the results substantially different from the pattern of the results of our prespecified analyses. Since the sensitivity analysis did not incorporate any indication about when the educational component of the intervention was delivered, it may be incorrect to characterize the observations during the first visits in 1998 or the observations in the first two visits in Group 2 as control, as is done in scenario one. In the absence of a protocol it is not possible to conduct a true intention-to-treat analysis.

We note that an effect, if present, on school attendance may not have arisen because of drug treatment. In a companion paper,[5] an effect reported in the original paper on the reduction in prevalence of anaemia was not present in re-analysis. In line with other research,[10] we found no short-term effect of the intervention on WAZ or HAZ in this statistical replication. These biological effects are key steps on causal pathways typically used to link deworming drug treatment with other benefits. Allocation to the intervention arm could therefore plausibly have affected school attendance through behavioural pathways affected by the educational component of the intervention, the placebo effect of the drug treatment or the Hawthorne effect.

Regarding generalizability of the intervention effect, worm burden needs to be high for schools to be eligible for the treatment. Burden may also affect the magnitude of effects: low burden may explain why a large trial in India evaluating the effect of deworming and vitamin A supplementation on pre-school mortality found no effect.[11] Without clear articulation of a causal pathway, it is unclear what other factors would need to be similar in other settings to generalize the results of this study.

This trial is, to our knowledge, the only published trial to investigate the effect of school-level deworming on educational outcomes. In our re-analysis, the strength of evidence that the deworming intervention improved school attendance was dependent on analytical choices, some of which are at risk of bias. The dataset had substantial amounts of missing data, hard-to-explain school attendance patterns and limited evidence for intermediate steps on the intervention's hypothesized causal pathway. We therefore conclude that this study provides some evidence for an effect of this complex intervention on school attendance, but with high risk of bias. For examination performance, there was no evidence of effect. We caution against generalizing these findings to other settings and recommend that further research is conducted.

The pre-analysis plan for this re-analysis can be found at [http://www.3ieimpact.org/media/filer/2013/05/14/ aiken_replication_plan_final.pdf].

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work was supported by the International Initiative for Impact Evaluation as part of the first round of their Replication Programme [http://www.3ieimpact.org/en/funding/replication-window/]. The aim of this programme is to improve the quality of evidence for development policy by re-appraising a wide range of influential studies in the development field, seeking to verify and examine the robustness of the original findings in these studies. The authors of this re-analysis report have conducted this work as consultants for 3ie. The funders have had no role in design, conduct or reporting of this re-analysis, other

than facilitating contact with the original study authors. Funding for the original intervention was provided by ICS Africa. Funding for the original evaluation of the effects of the intervention was provided by the World Bank Research Department, the Partnership for Child Development, the MacArthur Foundation and the University of California, Berkeley.

Acknowledgements

The replication would not have been possible without the cooperation and disclosure of the original authors. We would like to thank Professor Edward Miguel, Dr Joan Hamory Hicks, Michael Walker and Professor Michael Kremer for making available the original data for the study; for providing their analysis files; for providing detailed supporting documentation and for taking part in conference calls. We would like to thank Dr Ben Wood of 3ie for his support throughout the replication process and a number of reviewers associated with 3ie's replication programme for their helpful comments. This paper forms part of the work of the LSHTM Centre for Evaluation, which aims to improve the design and conduct of public health evaluations through the development, application and dissemination of rigorous methods, and to facilitate the use of robust evidence to inform policy and practice decisions.

Conflict of interest: We have no conflicts of interest to declare.

Note added in proof

Between first electronic publication and print publication, we identified various numerical errors in the manuscript and Appendix 5 as outlined below. The errors in Table 1, Table 2, and Table 4 were rounding errors or mistakes made when transcribing from the output of the analysis files. The errors in Table 3 and Appendix 5 were made due to mistakenly transposing equivalent analyses from a differently-defined study population. None of the errors led to any differences in interpretation and hence no textual changes were required. We notified the journal editors to these at the earliest opportunity and we apologise sincerely for these errors.

Main paper: Abstract. Year 2, adjusted logistic regression p-value corrected. Table 1. Mean age Group 2 lower CI corrected. Table 2. 1998, adjusted logistic regression (last column) p-value corrected. 1999, lower CI of difference in attendance corrected. Table 3. Difference in cluster summaries: 1998, Scenario One upper and lower CI corrected, and Scenario Two point estimate and upper and lower CIs corrected; 1999, primary analysis lower CI corrected, alternative Scenario One point estimate corrected, and Scenario Two point estimate and upper and lower CIs corrected. Adjusted logistic regression: 1998, Scenario One upper CI corrected; 1999, Scenario One and Scenario Two lower CI corrected. Table 4. Difference in proportion with moderate infection: round worm lower CI corrected, whip worm upper and lower CIs corrected. Weight for age Z-score: Group 1 and Groups 2+3 numbers tested (column 1) corrected. Height for age Z-score: Groups 2+3 number tested corrected, point estimate for difference between Groups 3 and 1 corrected, and upper CI for difference between Groups 3+2 and 1 corrected.

Appendix 5. Scenario One. Difference in cluster summaries: Year 1, lower and upper CIs corrected; Year 2, point estimate corrected. Logistic regression: Year 1, adjusted upper CI corrected; Year 2, unadjusted and adjusted lower CIs corrected. Scenario Two. Cluster summaries: Year 1 and 2, Groups 1,2, and 3 cluster summaries corrected. Difference in cluster summaries: Year 1 and 2, point estimates, upper and lower CIs, and p-values corrected. Logistic regression: Year 2, unadjusted and adjusted lower CIs corrected.

References

1. Bethony J, Brooker S, Albonico M *et al*. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *Lancet* 2006;**367:**1521–32.
2. Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect Dis* 2006;**6:**411–25.
3. Partnership for Child Development T. The cost of large-scale school health programmes which deliver anthelmintics to children in Ghana and Tanzania. *Acta Trop* 1999;**73:**183–204.
4. Miguel E, Kremer M. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 2004;**72:**159–217.
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. Begg C, Cho M, Eastwood S *et al*. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;**276:**637–39.
7. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol* 2011;**64:**936–48.
8. World Health Organization. *Helminth Control in School-Age Children: A Guide for Managers of Control Programmes*. Geneva: World Health Organization, 2002.
9. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007;**28:**182–91.
10. 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.
11. Awasthi S, Peto R, Read S *et al*. Population deworming every 6 months with albendazole in 1 million pre-school children in North India: DEVTA, a cluster-randomised trial. *Lancet* 2013; **381:**1478–86.