Breaking the cycle of infection: an impact evaluation of three strategies to control intestinal parasites and improve human capital in rural China

Linxiu Zhang, Chinese Academy of Sciences (CAS)

Renfu Luo, CAS

Alexis Medina, Stanford University

Chengfang Liu, CAS

Scott Rozelle, Stanford University

Xiaonong Zhou, China Centre for Disease Control and Prevention (CDC)

Yingdan Chen, China CDC

Grantee Final Report

Accepted by 3ie: June 2017

Note to readers

This impact evaluation has been submitted in partial fulfilment of the requirements of grant PW2.04 awarded under Policy Window 2. 3ie is making this final report version available to public as it was received from the authors. No futher copy-editing or formatting has been done. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was accepted. Any comments or queries should be directed to the corresponding author, Linxiu Zhang at lxzhang.ccap@igsnrr.ac.cn.

Suggested citation: Zhang, L, Luo, R, Medina, A, Liu, C, Rozelle, S, Zhou, X and Chen, Y, 2017, *Breaking the cycle of infection: an impact evaluation of three strategies to control intestinal parasites and improve human capital in rural China, 3ie Grantee Final Report.* New Delhi: International Initiative for Impact Evaluation (3ie)

Funding for this impact evaluation was provided by 3ie's donors, which include UK aid, the Bill & Melinda Gates Foundation and the Hewlett Foundation. A complete listing of all of 3ie's donors is available on the 3ie website.

Acknowledgements

We acknowledge financial support from the International Initiative for Impact Evaluation (3IE, grant number PW2.04.02.02). We are grateful to the thousands of public health officials and practitioners, school principals, teachers, village clinicians, and students who have helped us throughout this project.

Executive Summary

Soil-transmitted helminthes (STHs) infect more than one billion people around the world. There is concern that chronic infection with STHs among school-aged children may detrimentally affect children's development, including their health, cognition, and education. However, two recent Cochrane systematic reviews examining the impact of deworming drugs for STH on nutrition, hemoglobin and school performance found that randomized controlled trials in the literature provide an insufficient evidence base from which to draw reliable conclusions. This project uses a cluster-randomized controlled trial to assess the impact of a deworming + health education intervention on nutritional indicators, cognitive abilities and school performance among schoolchildren in rural China. The intervention, which was implemented by local health practitioners, consisted of a distribution of a 400 mg dose of albendazole accompanied with educational training about STH infection, treatment and prevention. The intervention was conducted twice during the duration of the study—at baseline in May 2013 and six months later in November 2013. We found that the deworming intervention significantly reduced both infection prevalence and infection intensity relative to the control group, but that these declines in infection were not accompanied by an impact on outcomes of nutritional indicators, cognitive abilities, or school performance. Our interpretation is that the impact of deworming was attenuated by the low levels of infection intensity in our sample population.

We conclude that if resources are not severely limited, deworming, which is a relatively inexpensive endeavor, will reduce STH infections and may help improve cognitive and educational indicators of children with high intensity STH infections. However, in settings with light infections, if fiscal resources are limited, other priorities may be chosen. Evidence from future RCTs is needed to assess the effect of deworming on key outcomes in other settings, especially those in areas in which the populations in infected with moderate and severe levels of worm infections. The evidence on effectiveness should also be combined with more precise data on costs of the intervention.

1. Introduction

Soil-transmitted helminths—*Ascaris lumbricoides, Trichuris trichiura, Necator americanus* and *Ancylostoma duodenale*—infect more than one billion people around the world.[1, 2, 3]. Studies have found that chronic infection with soil-transmitted helminths (STH) among school-aged children is associated with malnutrition and impaired growth,[4, 5, 6] cognitive impairment,[7, 8] and lower school attendance.[9] These associations suggest that, in theory, reducing STH infection in children has the potential to improve nutrition and growth (i.e. hemoglobin levels, weight, height), cognitive abilities (i.e. working memory, processing speed) and school performance (i.e. school attendance, standardized test scores).[10, 11] Recent studies have found evidence in support of long-run benefits of childhood deworming. [51, 52]

This causal model, however, has not been fully supported by the literature. A 2012 Cochrane systematic review examining the impact of deworming drugs for STH on nutrition, hemoglobin and school performance found that randomized controlled trials in the literature provide an insufficient evidence base from which to draw reliable conclusions.[12] The subsequent update of the systematic review by Cochrane in 2015 showed community deworming programs "probably have little effect on weight gain… and no effect on average cognition" with only 1361 participants in two trials with *low quality evidence*.[48] Furthermore, there is "probably no effect on height or …the average hemoglobin" looking at 3595 participants in seven trials with *low quality evidence*. Lastly there is "very limited evidence assessing an effect on school attendance and the findings are inconsistent and at risk for bias" (20,243 participants, in two trials and with *very low quality evidence*. In addition to the high risks of recruitment bias noted by the

authors of the Cochrane reviews, the current evidence base is limited by the following characteristics of existing studies: many trials are conducted with small sample sizes and are underpowered;[13, 14] many trials measure only one or two specific outcomes, often with selective reporting of outcomes;[12, 14, 15] and many trials do not report infection intensity,[12] which is measured by fecal egg counts, and which (if intensities are high or low) may have implications for the nature of the impact being measured. Furthermore, the majority of existing trials have been efficacy studies of individualised treatment, which tend to be researcher-implemented in a highly controlled setting.[16] The authors of the Cochrane report strongly recommend *effectiveness* studies (also known as pragmatic trials), which are ideally cluster-randomized controlled trials that examine the impact of the deworming intervention under real-world settings.[12, 17]

We designed a cluster-randomized controlled trial that is the first, to our knowledge, that involves a large sample size with high statistical power (>80% power); it examines the impact of deworming on a comprehensive list of nutrition, cognitive and school performance outcomes; and it reports intensity of STH infections based on measurements of fecal egg counts at baseline and follow-up. Our trial is an effectiveness study, rather than an efficacy trial, of a deworming program that was implemented by local health practitioners, thereby mirroring a real-world scenario that more pragmatically measures the impact of the intervention and may be a better way of informing policymaking.

China–the location of this study–currently lacks a regularized STH control program, despite historically moderate rates of STH prevalence in rural areas.[18, 19] For example, more than 40 percent of school-aged children in rural areas of Guizhou

Province in southwest China are infected with STH.[8, 18] According to Lu et al. (2015), there are three predominant reasons for high STH prevalence in rural areas of Guizhou province: lack of awareness and skepticism about STHs, local myths about STHs and deworming treatment, and poor quality of village health care.[49] The primary aim of this randomized controlled trial is to examine the impact of a deworming intervention on STH infection prevalence, infection intensity, nutritional indicators, cognitive abilities and school performance among school-aged children in rural Guizhou Province. In doing so, we also assess treatment compliance among children who were randomized to receive the intervention and offered treatment by public health officials.

2. Intervention, theory of change and research hypotheses

2.1 Intervention

The intervention consisted of a distribution of a 400 mg albendazole dose (two tablets of 200 mg) accompanied with two educational pamphlets (one for children and one for parents) about STH infection, treatment and prevention (*Figure 1*). Albendazole was manufactured by GlaxoSmithKline (GSK) and was purchased and shipped directly from the GSK warehouse to the county CDC. To mimic a real-world policy scenario, we consulted with health officials from the Chinese CDC to devise a plan for implementation. Health workers from local branches of the Chinese CDC were thus responsible for implementing the deworming intervention—distributing the albendazole and educational pamphlets—to children in the township schools randomized to receive the intervention. CDC health officials distributed albendazole in the classrooms twice during the duration of the study—at baseline in May 2013, and six months later in

November 2013—and instructed the children to take the tablets at home (Figure 2). Follow-up surveys and fecal sample collection were performed in April 2014. All children who were randomized to the control group of our study received albendazole after the conclusion of the study in April 2014.

2.2 Theory of Change and research hypotheses

We were also able to address external validity by examining the mechanisms underlying observed changes using causal chain analysis. We have mapped out the theories of change for the two different interventions, and present them below:

Systematic Deworming Intervention:

Correct Targeting of participants → Children aged 9-11 years old randomly sampled to participate the deworming + health education intervention → Intervention meets the project protocol → Deworming medication (Is the student receiving the deworming medication? Is the student receiving deworming medication of sufficient quality and quantity?) → Health education (Is the student attending the health training session? Is the health education material effective and applicable? Is his/her knowledge about STHs improving? Is his/her attitude towards STHS changing?) → Does the student take the deworming medication? Does he/she take sufficient quality / quantity of deworming medication? → Lower infection rates / better academic performance / better nutrition / better cognition/ better school performance.

Our survey instruments were designed to capture data from each step in the chain (i.e., whether the interventions were reaching the targeted population and/or having their intended benefits).

There could be several reasons why we might expect academic performance to be affected (or not be affected) by the interventions. Children who are free of intestinal worm infections might be less likely to suffer from nutritional deficiencies that could affect their concentration and school performance, or they might simply feel better and therefore have lower rates of absenteeism once infection levels are reduced. To capture these mechanisms, in addition to measuring pure academic performance, our survey instruments captured information on school attendance rates. We also conducted finger-prick blood tests for anemia, a common nutrient deficiency that has been shown both to be linked with intestinal worm infection and to have an impact on school performance.

If infection rates decrease, but there is no impact on either health indicators or educational performance, we used data from the control group to investigate the possibility that the lack of impact has its roots in teacher quality. It may be the case that children in the intervention groups are indeed better prepared to learn, but that the quality of schooling is so poor that the students are not actually taught anything. We consider this possible causal mechanism by measuring the real gain in learning in the control group. If the gain is extremely low, it may be the case that the impacts from the interventions are stymied by a poor learning environment.

In order to capture information on targeting in the deworming intervention, our student survey asked project participants whether they actually received the deworming medication they were supposed to receive, when they received it, and who else in the household (if anyone) actually took the medicine. This captures any lapses in the supply

chain, as well as any substitution away from the intended beneficiary of the implementation.

In order to control the quality of the deworming medication, we directly provided village clinicians with a standard, name-brand medication for distribution. To control for the quality of our outcome variables, we collected supplementary data on worm burden—the number of eggs per gram of stool. Even if actual prevalence rates do not fall significantly by the end of the project, it may be that the burden of infection has lessened, which has implications for the effectiveness of the medication and training.

In terms of academic performance, we used standardized tests carefully calibrated and extensively pretested in order to capture actual increases in performance, rather than simply test familiarity or natural increases in performance associated with an additional year of schooling. Based on previous studies run by the evaluation team, there is reason to believe that one year is sufficient time to capture changes in academic performance.

Finally, we sent a team of researchers trained in qualitative methodologies to the source villages where our sample children lived. These researchers conducted a qualitative survey of village health behaviors, discussing with villagers reasons for not adopting certain health behaviors (cost, perceived importance, inconvenience, etc.). They discussed reactions to the deworming medication, and attitudes towards worm infection in general, and towards village doctors in particular.

2.3. Research Hypotheses

a.) What is the impact of a two-prong approach (deworming program run through the public health practitioners + health education) intervention on student outcomes (prevalence and burden of STH infection, nutrition, cognition and school performance)?

b.) Why does the deworming + health education work or not work (causal chain analysis)?

c.) What kind of students do deworming + health education intervention work for and not work for (heterogeneous effects: migration status of parents, infection intensity, age, gender, ethnicity, socioeconomic status, academic achievement)?

3. Evaluation: Design, methods and implementation

3.1 Sample Size Determination

We conducted power calculations for our primary health and educational outcomes using "Optimal Design" software, developed by a group of researchers at the University of Michigan. To estimate the required sample size to observe reasonable effects on worm prevalence (our single binary outcome), we calibrated the model as appropriate for a cluster randomized trial with a binary outcome and clustering at the township level using parameters from a 2010 survey conducted as part of a previous project on intestinal worms in China. (For details about the math behind the calculation, please visit the Optimal Design website at http://sitemaker.umich.edu/group-based/optimal_design_software.)

The power needed to detect a difference in roundworm prevalence between the treatment and control groups in a cluster randomized trial with binary outcome depends on 6 factors:

a.) the number of children sampled per township (n);

b.) the number of townships (J);

c.) the township mean probability that a child is infected with any of the three types of worms (ascaris, hookworm, or whipworm) (ϕ_C);

d.) the lower plausible value of township mean probability (ϕ_{Clower});

e.) the upper plausible value of township mean probability (ϕ_{Cupper}); and,

f.) the township mean probability that a child who participates in the treatment will be infected with any of the three types of worms (ϕ_E).

Using results from a previous study of STH infections, we assume ϕ_C=0.34, ϕ_{Clower}=0.11, ϕ_{Cupper}=0.80. With 20 children per township (20 fourth and fifth grade students – see sampling below), we estimate that 55 townships per treatment arm (110 townships per inter-arm comparison; township villages total) will provide 80% power to detect a difference in infection prevalence between any two arms of 0.11 or greater (or, ϕ_E=.23).

Besides the probability of worm infection, all of the other primary outcomes in the study are continuous variables. These include worm burden, hemoglobin concentration, cognitive testing (TIMSS test for children aged 9-11 years), scores on the WISC test, and child weight-for-age, height-for-age, and BMI-for-age.

For these outcomes, we conducted power calculations using the same software as for infection incidence (Optimal Design: http://sitemaker.umich.edu/group-based/optimal_design_software). Calculating power for continuous outcomes in a cluster randomized design requires the following assumptions:

a.) the number of children per township (n);

b.) the number of townships (J);

c.) the intra-township correlation in outcome variable (ρ);

d.) the minimum effect size that we would expect to be able to detect (δ); and

e.) the proportion of variation in the true township-level post-intervention outcome variable explained by township-level pre-intervention outcome variable (R^2).

Appendix Table 1 below gives results for our continuous outcomes given varying effect sizes. Assumptions on ρ and R2 are based on our own data from previous studies. For health outcomes, we estimate power for the pooled sample. In the table, the next to last column gives the calculated power at 110 townships per comparison (55 townships per arm) and the last column gives the needed number of townships per comparison (divide by 2 for the number per arm) for 80% power. Note that these calculations are somewhat conservative given potentially substantial gains in power from ex-ante matching of townships in treatment assignment. [29,30]

3.2 Power Calculations for Subgroup Analyses

Our analysis of heterogeneous treatment effects focuses on estimating treatment effects on health outcomes. Appendix Table 2 below gives our calculations for each of the (continuous) health outcomes. The parameters for all of these outcomes are again taken from our previous worm prevalence survey which used a similar sampling strategy. The last two columns, as above, give the calculated power at 110 townships per comparison (55 schools per arm) and the needed number of townships per comparison for 80% power.

Our power to detect reasonable effect sizes – 0.25 for worm infection and hemoglobin outcomes and 0.2 for anthropometrics – is above 80%.

For worm prevalence, we recalculate power using the same (binary outcome) method as above. We assume that ϕ_C =0.44, ϕ_{Clower} =0.2, and ϕ_{Cupper} =0.9. With these parameters, a sample of 110 villages per comparison and 20 fourth and fifth graders per township will provide enough power to detect a difference in infection prevalence between any two arms of 0.14 or greater (or, ϕ_E =0.3) for fourth and fifth graders.

4. Context and Sampling Design

4.1 Context

The study was conducted in Guizhou Province, located on the eastern part of the Yunnan-Guizhou Plateau in southwest China. Guizhou's population, estimated to be about 35 million in 2011, is demographically diverse with ethnic minority groups making up more than 37 percent of the population [27] Guizhou Province is the second poorest of China's 31 provinces, with most people living in extremely remote mountainous areas that lack adequate road infrastructure. Agriculture is the main occupation of most Guizhou inhabitants, and farmers in Guizhou often live in poverty with poor access to basic health services.[27]

The warm, humid climate in Guizhou, combined with poor sanitation in local villages, offers ideal conditions for worms to breed, thrive, and spread. Our previous survey work in this area has shown that the average prevalence among children in Guizhou is around 40 percent. In some villages, the prevalence is as high as 80 percent.

3.2 Sampling Design

In accordance with the power calculations, this cluster-randomized controlled trial included a total of 2,240 sample children and spanned 112 townships in seven of the poorest rural counties in Qiandongnan Prefecture in Guizhou Province (*Figure 3*). The

seven counties were randomly selected from the poorest half of counties in Qiandongnan prefecture based on per capita income, according to figures published by the Guizhou Provincial Bureau of Statistics. [31] According to national statistics, at 4,625 yuan, the average rural individual in our sample areas has a per capita income in the bottom quartile of China's rural income distribution.[27]

All townships within each of the selected counties were included in our sample, except for those which housed the county government; these townships were excluded because they are generally wealthier and more urbanized than the average rural township. A total of 112 townships met our selection criteria.

In each township, we obtained from the central primary school a roster of all children aged 9 to 11 years old attending any primary schools within the township for the 2013-2014 school year. We focused on this age group because school-aged children typically have the highest burden of STH infection,[28] and specifically, elementary school children aged 9 to 11 years in our study area are old enough to take standardized exams. We classified all 9 to 11 years old children by their home village, then randomly selected 20 sample children from the home village with the largest number of children at that school. We excluded villages that housed the local township government, since these villages are typically wealthier and more urban than a typical village. If the first home village we selected had fewer than 20 children in our age group attending the school, we randomly selected children from the next-largest village to fill in the gap. In total, then, 20 school children from either one or two villages in each township were randomly chosen for participation in the study. Overall, our sample

population was composed of 2,240 children from 146 villages in 112 townships in 7 rural counties.

As it turned out, a total of 2,179 students were enrolled in our study at baseline in April 2013: 1,084 children in the intervention group and 1,095 children in the control group. Of the 2,179 students enrolled, 151 were lost to follow-up in May 2014: 84 students from the intervention group and 67 from the control group (Figure 3). Attrition was due either to students transferring to other schools or to missing fecal sample information. A total of 2,028 participants (93% of the enrolled sample) were included in the follow-up analysis: 1,000 children in the intervention group and 1,028 children in the control group.

5. Randomisation and masking

Cluster randomisation was conducted at the township level. All randomized selection and allocation was performed using a computerized random sequence generator. In each of the seven counties included in our study, we randomly assigned half of the townships within each county to the control group and the other half to the intervention group. To increase statistical power, we used baseline survey information to assign the sample townships in each county into two pairs, using an optimal matching algorithm. The optimal matching algorithm assigned sample townships into pairs by minimizing the total (Mahalanobis) distance within the matched pairs.[32] The Mahalanobis distance measure was calculated using the following baseline covariates at the township level: prevalence of STH, per capita net income, prevalence of anemia,

number of households with children between the ages of 3 and 18 years, and distance (km) to the nearest paved road.

After matching sample townships into pairs, we randomly assigned one township in each pair to either a control or intervention group. In our study, 56 out of a total of 112 townships were randomly assigned to receive the intervention (intervention group). The remaining 56 townships were assigned to the control group, which did not receive the intervention. The risk of spillover effects was low, given that paired townships were separated by more than 50 minutes of driving time, on average. In addition, no two schools were in the same school district.

Trained enumerators and local health practitioners who assisted with baseline and follow-up surveys were not explicitly informed of the treatment assignment of participants, although blinding of participants themselves was not possible because of the nature of the intervention. Students in the intervention group, as well as their parents or teachers, were not told explicitly that the purpose of the study was to examine the effect of a trial intervention. The study team informed students that they were participating in a general study of health and education of rural pupils by the Chinese Academy of Sciences and the Chinese Centers for Disease Control and Prevention (CDC). Participants in the control group were not aware that they were in a randomized trial.

6. Ethics

This study received ethical approvals from the Stanford University Institutional Review Board (IRB) (Protocol ID 25027), and from the Sichuan University Ethical

Review Board (Protocol ID 2013005-02). All participating children gave oral assent prior to baseline data collection, and the children's legal guardians gave written consent for their children's involvement in the study. Children who were found to have severe anemia were referred to the local hospital for treatment. All participants were provided with deworming medication at the conclusion of the study.

7. Procedures

Baseline surveying and fecal sample collection were performed in early May 2013. For each participating student, we obtained fecal samples for parasitological testing, administered a socioeconomic survey regarding individual and family characteristics, performed a physical exam to obtain measures of nutritional indices, and conducted standardized tests to assess cognitive abilities and school performance.

For parasitological testing, the study team collected two fecal samples from each child in our sample: one fecal sample per day for two consecutive days. Samples were picked up once per day by the study team and were stored in a temperature-controlled cooler until collection. At the time of collection, members of the study team transported all fecal samples in a temperature-controlled cooler to the laboratory of the county branch of the CDC. A total of 2,179 children who produced at least one stool sample were included in our analysis. All fecal samples were tested on the same day that they were collected. Fecal samples were analyzed microscopically at the county CDC laboratory using the Kato-Katz thick-smear technique for *Ascaris lumbricoides* (*Ascaris*), *Trichuris trichuria* (*Trichuris*), and *Ancylostoma duodenale* or *Necator americanus* (hookworm). [34] Two smears were taken from each of the two fecal samples collected

from each child: one smear from each of the two samples was tested the same day on-site. The second smear from each sample was treated using a formaldehyde preservation technique and sent to the headquarters of the National Institute for Parasitic Diseases in Shanghai for quality control analysis and to perform egg counts for intensity of infection. Children were considered positive for STH infection if at least one of their fecal samples tested positive for one or more species of STH. Among fecal samples that tested positive for STH, we calculated fecal egg count by quantifying the geometric mean number of eggs per gram of feces in each sample. Categorization of infection intensity as light, moderate, or severe was assigned according to World Health Organization classification, based on mean fecal egg count and STH species.[35]

The socioeconomic survey consisted of questions regarding the demographic characteristics and household conditions of children and parents. Students completed the survey in their classrooms under the supervision of trained enumerators from the Chinese Academy of Sciences and Guizhou University of Finances and Economics.

The physical exam measured three nutritional indicators: hemoglobin (Hb) concentrations, height and weight. Hb levels were measured using HemoCue Hb 201+ systems. Height and weight measurements were obtained following WHO standard protocol. [37] The children were measured in light clothing without shoes, hats or accessories. Weight was measured with a calibrated electronic scale recommended by professionals from the West China School of Public Health of Sichuan University. Body height was measured using a standard tape measure. The nursing team was trained to set up the weighing station on level ground to ensure accuracy of the equipment. Two nurses manned each measurement station, with one responsible for preparing subjects

for measurement (removing shoes, offering instruction, positioning children, etc.) and the other responsible for conducting and recording the measurements.

Cognitive ability was assessed using a battery of four sub-tests from the Mandarin-language version of the latest Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV) (*Appendix Table 3*). The WISC-IV tests were culturally adapted, translated and edited into simplified Chinese and validated for assessment among Chinese children in 2008. [37] According to the literature, children's working memory and processing speed are the cognitive areas that are most likely to be affected by STH infection; [38, 39] thus, we focused our efforts on measuring these two outcomes. In the WISC-IV, working memory index (WMI) is assessed through two core subtests: Digit Span and Letter Number Sequencing. Processing Speed Index (PSI) is also assessed through two core subtests: Coding and Symbol Search. Trained examiners administered these four core subtests of cognitive ability to all children participating in the study on a one by one basis.

Measures of school performance included attendance rates and scores on the Trends in International Mathematics and Science Study (TIMSS), an internationally-utilized standardized test established by the International Association for the Evaluation of Educational Achievement to compare student educational achievement internationally.[40] School attendance rates were obtained from reports recorded by homeroom teachers.

8. Outcome variables and Statistical Analysis

9*.1 Outcomes*

The primary outcomes analyzed were STH infection prevalence, stunting prevalence (HAZ < -2), underweight prevalence (WAZ < -2), working memory index (WMI), processing speed index (PSI), school attendance and normalized TIMSS mathematics test scores. Secondary outcomes analyzed were infection intensity (fecal egg counts) and anemia prevalence.

Measurements of hemoglobin levels were used to determine anemia prevalence. Following WHO guidelines, anemia is defined as having a hemoglobin level of less than 115 g/L for children 5 to 11 years of age, and less than 120 g/L for children 12 to 13 years of age.[20] Measurements of height and weight were used to construct body mass index (BMI)-for-age z-scores and height-for-age z-scores (HAZ) using WHO AnthroPlus, a software application of the WHO Reference 2007 for children aged 5-19 years that is used to monitor the growth of school-aged children and adolescents.[21] Weight-for-age z-scores (WAZ) were calculated using a SAS program for the 2000 U.S. CDC growth chart for children aged 0-20 years.[22] Raw scores obtained from the core sub-tests of the WISC-IV were converted to age-scaled index scores using the tables of norms in the Mandarin version of the WISC-IV administration and scoring manual to produce the index scores for Working Memory Index (WMI) and Processing Speed Index (PSI) that were analyzed in this study. Scores on the Trends in International Mathematics and Science Study (TIMSS) were normalized by the distribution of the control group in both the baseline and the follow-up surveys, and school attendance reports were used to calculate attendance rates.

Treatment compliance rates were also analyzed. Following each round of deworming treatment in May 2013 and November 2013, students in the intervention

group were asked to fill out a brief survey regarding the number of albendazole pills that they took (zero, one, or two). Treatment compliance rates were obtained from the responses of students.

9.2 Statistical analysis

Among the 10 outcomes of interest, the largest sample required to meet at least a 0.25 standardized effect was for worm prevalence, based on parameters from previous studies. With a sample of 100 townships (50 control, 50 treatment) and 20 children per township, we estimated a 12% decline in worm prevalence at 80% power. We increased the sample size to 112 townships to account for potential attrition. We assumed a prevalence in the control group of 34% with a 95% plausible interval of 11% to 80%. Power calculations were performed with Optimal Design software from the University of Michigan using the option for a cluster-randomized trial with a binary outcome. Our sample size provided adequate power to detect meaningful effects on the other 11 outcomes of interest. Details are available from the principal investigator upon request.

To further increase the power of the trial, we used a pairwise matching randomization procedure (as discussed in the “randomization and masking” section above). While not explicitly accounted for in determining the required sample size, the power gains from matching are potentially substantial.[23, 24]

All statistical analyses were performed using STATA 12.0 (STATA Corp.; College Station, United States of America). P-values below 0.05 were considered statistically significant. We report coefficients and 95% confidence intervals (CIs). Comparisons between the intervention and control groups for all outcomes by subgroup populations

were assessed using a t-test. Multivariate analyses for the continuous outcome measures—fecal egg count, HAZ, WAZ, WMI, PSI, as well as the normalized TIMSS mathematics test scores—were performed using STATA's multiple linear regression model and its estimation using ordinary least squares (OLS), taking into account the pairing nature of townships within county and data-clustering at the township level. Multivariate analyses of binary outcome measures—STH prevalence, anemia prevalence, and school attendance—were performed using STATA's logistic regression model, also taking into account paired fixed effects and clustering at the township level. Following previous studies,[8, 18, 25] we adjusted for the following two sets of additional covariates at baseline survey in the multivariate analyses to increase statistical precision: student individual characteristics (gender; age; boarding status; belonging to the Dong, Miao, or Shui minority groups) and household characteristics (number of siblings, pieces of durable assets, migration status of parents, educational attainment of the parents). We also included pair fixed effects at the township level. All p-values were based on results from the adjusted model.

We supplemented our intention to treat (ITT) multivariable analyses (described above) by examining the average-treatment-effects-on-the-treated (ATT analysis) to measure the impact on outcomes among the subpopulation of children who were fully compliant with treatment, thereby controlling for any confounding due to non-compliance. For ATT analysis, we used an instrumental variable (IV) approach in which the treatment assignment was used to instrument for observed compliance,[26] thereby allowing us to measure the effect of treatment among the subpopulation of children who reported full compliance with treatment, and thus control for confounding due to non-

compliance. ATT analyses for the continuous outcome measures were performed using STATA's ivreg model, for the binary outcome measures using STATA's ivprobit model. Both models take into account the pairing nature of townships within county and data-clustering at the township level.

Additional analysis of the correlation between infection intensity and outcomes was determined by calculating pairwise correlation coefficients between fecal egg count and outcome measures at the baseline survey among samples with positive infection.

The trial was registered with the ISRCTN Registry in April 2013 (trial number: ISRCTN97311712), prior to the start of study activities.

9. Program or policy: Design, methods and implementation

9.1 Analysis Design

This evaluation was run by academics and policy analysts in the Chinese Academy of Sciences; the project was implemented by officials, medical professionals and scientists in the Chinese Center for Disease Control; and the intended beneficiaries are Chinese children, their families, and health care professionals in townships in Southwest China.

Specifically, the project evaluators at the Chinese Academy of Sciences (CAS) were trained as economists and specialize in impact evaluations. To date they have conducted 26 Randomized Controlled Trials in which they have been the lead PI. The CAS team were responsible for designing, coordinating, and implementing the project evaluation from start to finish. Their position as part of a state-run research institution allows them access to policy channels unavailable to researchers at universities. The

CAS evaluation team are able to put together an official “policy brief” distilling our results into a four-page document with a cogent, resonant message. This brief can then be forwarded through the Chinese Academy of Sciences to the policy advice desk at the Office of the Premier.

The main project implementer is the China Center for Disease Control & Prevention (CDC), a nationwide organization in China that is charged with a multitude of tasks, including organizing and implementing disease control and prevention plans, providing staff training and quality control for disease prevention activities, and implementing public health projects for women and children. Among other things, the CDC is in charge of controlling infectious diseases, providing and keeping records of immunizations and providing emergency relief services in times of national crisis. Headquartered in Beijing, there are CDC units and assigned personnel in each and every level of the government hierarchy (central government; provincial government; prefecture; county; township).

The National Institute for Parasitic Disease (NIPD) is a standalone division of the national CDC. Its stated mission is to track and control the prevalence of parasitic diseases across China. Its headquarters are in Shanghai (for historic reasons). Historically most of its efforts were focused on eliminating schistosomiasis. NIPD has also been fully engaged in campaigns against malaria, filariasis, and echinococcosis. The past experience, current regular duties and authority over provincial and county CDCs of the NIPD makes their team fully qualified to direct the implementation of this project. In the recent past they have run projects to improve latrines in rural areas and to educate the populace on the transmission of intestinal worm infections. Currently,

they are involved in a program that measures the effect of a new cultural training program to instill good hygiene behaviors.

9.2 Monitoring System

The past experience, current regular duties and authority over provincial and county CDCs of the NIPD makes their team fully qualified to direct the implementation of this program. In the recent past they have run projects to improve latrines in rural areas and to educate the populace on the transmission of intestinal worm infections. Currently, they are involved in a program that measures the effect of a new cultural training program to instill good hygiene behaviors.

9.3 Recruitment Strategy

No recruitment was conducted as part of this study. Instead, sample schools and sample students within those schools were randomly selected as per our sampling strategy described above (see section 5.4). A total of 2,179 students were enrolled in our study at baseline in April 2013: 1,084 children in the intervention group and 1,095 children in the control group. Of the 2,179 students enrolled, 151 were lost to follow-up in May 2014: 84 students from the intervention group and 67 from the control group (Figure 3). Attrition was due either to students transferring to other schools or to missing fecal sample information. The intervention, which was implemented by local health practitioners, consisted distributing a 400 mg dose of albdenazole, accompanied with educational training about STH infection, treatment and prevention.

9.4 Implementation Weaknesses

Regarding infection prevalence, reinfection dynamics often cause levels of STH prevalence to persist in a population, a factor that should be considered when

evaluating the impact of future deworming interventions. The pattern of rapid reinfection after antihelminthic treatment (in our case we tested the follow-up fecal samples 6 months after treatment) has been consistently identified in geographically diverse populations in both children and adults, [41] and was previously confirmed in our study area.[8, 18] In our scenario, we surmise that children in the intervention group were cleared of STH infection after each of the two rounds of antihelminthic treatment, but experienced high rates of reinfection within the subsequent months leading up to post-intervention evaluation. In a 2012 article by Jia et al., helminth reinfections occur rapidly after treatment, particularly for *A. lumbricoides* and *T. trichiura*. [42] Hence, there is a need for frequent anthelmintic drug administrations to maximize the benefit of preventive chemotherapy. Integrated control approaches emphasizing health education and environmental sanitation are needed to interrupt transmission of STH. Bieri et al. (2013) found that a cartoon video health education package increased student knowledge about STHs and led to change in behavior and a reduced incidence of infection within one school year.[50] Additionally, natural fluctuations in infection in the environment, as previously observed in other populations, [42, 43, 44] may have caused the decrease in infection prevalence in the control group and further attenuated the between-group difference in prevalence reduction.

Additionally, the majority of fecal samples from the children were not produced on-site at the time of collection. Thus, it is possible that there was a delay of up to a few hours before children delivered their samples to refrigeration facilities at the school or village clinic. This may have caused an underestimate of total STH prevalence, especially with respect to hookworm; therefore, estimates presented in this report can

be considered a lower bound for actual infection prevalence and infection intensity in our study population. In addition, our study focused on examining short-run impacts of deworming; studies are also needed that follow children with no STH infections for longer intervals of time compared to controls with higher (any) levels of infection intensities.

10. Impact analysis and results of the key evaluation questions

10.1 Balance between the treatment and control groups at the baseline

All of the study outcomes were balanced between the control and intervention groups at baseline, as were nearly all of the control variables (Table 2).

In our survey, we took caution to ensure the quality of data collection and implementation of the intervention. As a result, in our analyses, we did not see any extreme outliers that were distant from other observations in terms of our outcome variables. Therefore, we did not exclude any outliers from our analyses.

Some outcomes were self-reported. Students completed the socioeconomic survey of questions regarding the demographic characteristics and household conditions of children and parents in their classrooms under the supervision of trained enumerators from the Chinese Academy of Sciences and Guizhou University of Finances and Economics. Additionally, following each round of deworming treatment in May 2013 and November 2013, students in the intervention group were asked to fill out a brief survey regarding the number of albendazole pills that they took (zero, one, or two). Treatment compliance rates were obtained from the responses of students.

The groups were statistically identical on all outcome measures at the time of the baseline survey (*Table 2*). In the control group, 30.50% were infected with *Ascaris,* 23.29% with *Trichuris,* and 1.00% with hookworm; 12.97% were co-infected with *Ascaris* and *Trichuris*. In the intervention group, 31.09% were infected with *Ascaris,* 24.35% with *Trichuris,* and 0.74% with hookworm; 12.92% were co-infected with *Ascaris* and *Trichuris*.

10.2 Results from Analysis Based on Intention to Treat

10.2.1 Prevalence and intensity of infection

Table **3** compares infection prevalence and infection intensity for the intervention and control groups from baseline to follow-up. At baseline, the prevalence of STH infection was 41.10% (95% CI, 38.18 to 44.01%) in the control group and 42.62% (95% CI, 39.67 to 45.57%) in the intervention group (odds ratio in the intervention group, adjusted for individual and household characteristics, 1.15; 95% CI, 0.93 to 1.43). There was no significant between-group difference in infection prevalence at baseline (p=0.192). At follow-up, the prevalence of any STH infection was 31.40% (95% CI, 28.56 to 34.23%) in the control group and 27.66% (95% CI, 24.89 to 30.43%) in the intervention group (adjusted odds ratio in the intervention schools, 0.71; 95% CI, 0.52 to 0.96). There was a significant between-group (treatment vs. control) difference in infection prevalence at follow-up (p=0.026).

At baseline, the mean fecal egg count, assessed as the geometric mean number of eggs per gram (epg) of feces among positive-tested samples, was 493.68 epg (95% CI, 357.83 to 629.53) in the control group and 702.71 epg (95% CI, 491.07 to 914.34) in the intervention group. According to World Health Organization (WHO) categorization,

both groups had light-intensity infection, defined as a mean egg count of 1 to 4,999 epg for *Ascaris* and 1 to 999 epg for *Trichuris*.[45] There was no significant between-group difference in mean fecal egg count at baseline (p=0.293). At follow-up, the intensity of STH infection was 533.32 epg (95% CI, 390.69 to 675.95) in the control group and 299.93 epg (95% CI, 193.98 to 405.89) in the intervention group. STH infection intensity from baseline to follow-up increased by 39.64 epg in the control group, whereas it decreased by 402.78 epg in the intervention group. There was a significant between-group difference in infection intensity at follow-up (p=0.018).

10.2.2 Nutritional indicators

Table 4 compares nutritional indicators, cognitive abilities and school performance for the intervention and control groups from baseline to follow-up. At baseline, the mean hemoglobin level was 126.17 g/L in the control group (95% CI, 125.43 to 126.91) and 126.33 g/L in the intervention group (95% CI, 125.59 to 127.07); there was no significant between-group difference at baseline (p=0.604). At follow-up, the mean hemoglobin level was 132.16 g/L in the control group (95% CI, 131.37 to 132.95) and 131.73 g/L in the intervention group (95% CI, 130.92 to 132.54). Hemoglobin levels increased among children in both the control and intervention groups. There was no significant between-group difference at follow-up (p=0.623).

At baseline, the anemia prevalence was 16.62% in the control group (95% CI, 14.41 to 18.83%) and 16.14% in the intervention group (95% CI, 13.95 to 18.34%); there was no significant between-group difference at baseline (p=0.385). At follow-up, the anemia prevalence was 9.98% in the control group (95% CI, 8.14 to 11.82%) and

11.58% in the intervention group (95% CI, 9.61 to 13.55%). There was no significant between-group difference at follow-up (p=0.174).

At baseline, the prevalence of stunting, defined as HAZ < -2, was 26.98% (95% CI, 24.32 to 29.65%) in the control group and 29.66% (95% CI, 26.89 to 32.42%) in the intervention group. There was no significant between-group difference in the prevalence of stunting at baseline (p=0.291). At follow-up, 23.48% (95% CI, 20.85 to 26.11%) of children in the control group and 27.63% (95% CI, 24.84 to 30.42%) of children in the intervention group were stunted. There was no significant between-group difference at follow-up (p=0.367).

At baseline, the prevalence of children who were underweight, defined as WAZ < -2, was 24.11% (95% CI, 21.57 to 26.65%) in the control group and 28.90% (95% CI, 26.20 to 31.61%) in the intervention group. There was a significant between-group difference in the underweight prevalence at baseline (p=0.004). At follow-up, 21.37% (95% CI, 18.85 to 23.88%) of children in the control group and 24.19% (95% CI, 21.56 to 26.82%) in the intervention group were underweight. There was no significant between-group difference at follow-up (p=0.113).

10.2.3 Cognitive abilities

At baseline, mean Processing Speed Index (PSI) score was 86.21 points (95% CI, 85.44 to 86.99 points) in the control group and 86.09 points (95% CI, 85.31 to 86.87 points) in the intervention group. There was no significant between-group difference at baseline (p=0.827). At follow-up, the mean PSI score was 88.18 points (95% CI, 87.37 to 88.99 points) in the control group and 88.83 points (95% CI, 88.01 to 89.65) in the

intervention group. There was no significant between-group difference at follow-up (p=0.143).

At baseline, mean Working Memory Index (WMI) score was 78.68 points (95% CI, 78.08 to 79.27 points) in the control group and 78.51 points (95% CI, 77.92 to 79.10 points) in the intervention group. There was no significant between-group difference at baseline (p=0.922). At follow-up, mean WMI score was 78.23 points (95% CI, 77.61 to 78.86 points) in the control group and 78.50 (95% CI, 77.86 to 79.14 points) in the intervention group. There was no significant between-group difference at follow-up (p=0.093).

10.2.4 School performance

At baseline, the school attendance rate was 86.73% (95% CI, 84.65 to 88.81%) in the control group and 87.32% (95% CI, 85.29 to 89.35%) in the intervention group. There was no significant between-group difference at baseline (p=0.692). At follow-up, the school attendance rate was 86.13% (95% CI, 83.81 to 88.45%) in the control group and 85.30% (95% CI, 82.95 to 87.66%) in the intervention group. There was no significant between-group difference at follow-up (p=0.496).

At baseline, the normalized score on the TIMSS math assessment was 0.00 (by normalization) in the control group (95% CI, -0.06 to 0.06) and -0.04 in the intervention group (95% CI, -0.10 to 0.02). There was no significant between-group difference at baseline (p=0.912). At follow-up, the normalized score on the TIMSS math assessment was 0.00 (by normalization) in the control group (95% CI, -0.06 to 0.06) and -0.07 in the intervention group (95% CI, -0.14 to -0.01). There was no significant between-group difference at follow-up (p=0.190).

10.3 Results from Analysis Based on The Average Treatment Effect on the Treated

Some causal chain analysis was conducted by assessing the treatment compliance among children in the intervention group after both the first and second rounds of deworming. Average-treatment-effect-on-treated analysis was conducted as well, whose results mirrored those of the primary intention to treat (ITT) analysis of the full sample of children.

10.3.1 Treatment compliance

Treatment compliance was assessed among children in the intervention group after both the first and second rounds of deworming (May 2013 and November 2014, respectively). Out of 1000 sample children in the intervention group who were included in the follow-up analysis, 52.21% reported taking the complete dose of two 200 mg albendazole pills in both rounds of deworming, and 75.92% reported taking at least one of the two 200 mg albendazole pills in both rounds. When assessing compliance survey responses in the second round of deworming in November 2013 only, we found that 63.37% of children took the complete dose of two 200 mg albendazole pills and 82.10% took at least one of the two pills. Detailed information about the distribution of the number of albendazole pills taken is presented in Appendix Table 5.

10.3.2 Results from the Average-treatment-effect-on-treated analysis

The results of the ATT analysis mirrored those of the primary intention to treat (ITT) analysis of the full sample of children: the intervention had a significant impact on reducing STH infection prevalence and infection intensity, but no evidence of an effect

on any of the other measured outcomes. The ATT analysis found that children who reported being compliant with deworming treatment experienced significantly greater reductions in both infection prevalence ($p = 0.011$) and infection intensity ($p = 0.019$) (Appendix Table 6). The point estimates generated by the ATT analysis were greater than those generated by the primary ITT analysis (-0.39 versus -0.21 for infection prevalence, and -370.37 versus -209.78 for infection intensity). Importantly, the greater impact of the ATT analysis indicates that our compliance variable is at least in part accurately measuring student behavior. As with the ITT analysis, there was no evidence of a significant impact of deworming on the measured outcomes of hemoglobin levels ($p = 0.622$), anemia prevalence ($p = 0.185$), stunting prevalence ($p = 0.335$), underweight prevalence ($p = 0.174$), Processing Speed Index score ($p = 0.142$), Working Memory Index score ($p = 0.093$), school attendance rate ($p = 0.491$), or normalized TIMSS score ($p = 0.187$).

10.4 Correlation between infection intensity and primary outcomes

We conducted an additional analysis to assess the correlation between infection intensity and primary outcome variables among participants in our sample at baseline (*Table* 5). We find that there was a strong correlation ($p<0.05$) between fecal egg counts and outcomes of cognition and school performance. Higher fecal egg counts, indicating more severe infection intensity, was associated with lower Processing Speed Index ($R = -0.12$, $p = 0.007$), lower Working Memory Index ($R = -0.13$, $p = 0.004$), and lower TIMSS mathematics test scores ($R = -0.14$, $p = 0.002$).

No significant heterogeneous effects were observed by subgroups in terms of age, gender, ethnicity or social economic status.

10.5 Heterogeneous Impacts

We conducted an additional analysis to assess the correlation between infection intensity and primary outcome variables among participants in our sample at baseline (*Table 5*). We find that there was a strong correlation ($p<0.05$) between fecal egg counts and outcomes of cognition and school performance. Higher fecal egg counts, indicating more severe infection intensity, was associated with lower Processing Speed Index ($R = -0.12$, $p = 0.007$), lower Working Memory Index ($R = -0.13$, $p = 0.004$), and lower TIMSS mathematics test scores ($R = -0.14$, $p = 0.002$).

While our evidence shows that the intervention had no impact on nutritional indicators, cognitive abilities, or school performance in this lightly-infected population, the impact on populations with moderate-to-high levels of infection intensity is an area for further investigation. Our additional descriptive analysis identified a significant correlation between higher levels of infection intensity and worse measures on key outcomes of cognition and school performance (*Table 5*).

We also conducted sub-group regression analysis based on worm load intensity. Since a majority of our infected sample children belong to the low infection intensity by the WHO classification, we made our working grouping. Specifically, we grouped those infected children into relatively lower infection intensity whose worm loads are less than 1000 epg, and those who have relatively higher worm intensity (epg<=1000). Then, we reran the adjusted model with the relatively lower group, and the relatively higher group, respectively.

Our results from the sub-group analyses show that for the relatively lower worm load group, the intervention has significantly impact infection intensity, which is consistent with the results from analyses using the whole sample (Appendix Table 3). However, the intervention also has significant impact on Hb level, anemia prevalence, underweight prevalence, and standardized math test score. For the relatively higher worm load group, the intervention has significant impact on stunting rate and working memory index score, but not on other outcome variables.

In addition, we conducted heterogeneous effect analysis by interacting the treatment dummy with the dummy indicating whether both parents outmigrate. Our data show that if treatment students come from households with both parents absent, they tend to have lower probability of attending school, and poorer academic performance as measured by standardized math test scores (Appendix Table 4).

10.6 Cost-Benefit Analysis

Since we find no empirical evidence that the intervention can be justified on the basis of improvement of nutritional indicators, cognitive abilities or school performance, should the Chinese government allocate resources for deworming in areas with light-intensity STH infection? If resources are not severely limited, deworming, which is a relatively inexpensive endeavor (at USD 0.65/year/child), will reduce STH infections and, according to the correlations, may help improve cognitive and educational indicators of children with high intensity STH infections (even though the degree of rise may be too small to detect in a study of the scale of this current study)[1]. However, the results of this

[1] The team has calculated the unit cost number based on price information from the Jingdong (https://item.jd.com/10741403238.html#crumb-wrap), one of the largest B2C online retailers in China. The

effectiveness study indicate that the intervention is unlikely to produce a rapid impact on nutritional outcomes. Thus, if resources are scarce, public health efforts might best be concentrated on other interventions—though the final decision ought to be supported by detailed cost-benefit analysis.

11. Discussion

Our cluster-randomized controlled trial assessed the impact of a deworming intervention on nutritional indicators, cognitive abilities and school performance. Our results show that the intervention significantly reduced both infection prevalence and infection intensity relative to the control group. These declines in infection however, were not accompanied by an impact on outcomes of nutritional indicators, cognitive abilities, or school performance. These results are consistent with the most recent replication studies by Aiken et al. (2015) in that although deworming reduces STH infections, it has no impact on anemia prevalence, school attendance or academic performance. Davey et al. (2015) also concluded that cautions needs to be taken when interpreting the results that a school-based drug-treatment and health-education intervention improved school attendance and no evidence of effect on academic performance.

The most recent Cochrane review (July 2015) about treating school children for worms states: “In trials that treat only children known to be infected, deworming drugs may increase weight gain (low quality evidence), but we do not know if there is an effect on cognitive functioning or physical well-being (very low quality evidence).” Our study

price for a 10-tablet of 200 mgs of packet is about 11 RMB yuan, in the study area, if we deworm a child twice a year with a dosage of 400 mgs each time, the cost will be (11/10*2*2) = 4.4 yuan, which is equivalent to about 0.65 USD at an exchange rate of 6.8 yuan per USD.

adds to the quality of evidence to help make informed decisions about treating children in helminth-endemic areas. The design (RC cluster), and measures (growth and cognitive impacts) of our study address some of the key shortcomings delineated in the 2012 and 2015 Cochrane Deworming reviews which cite a need for more and better quality (GRADE) studies with randomized cluster design. While there was no observed impact in our study on any of the nutrition, cognition or educational outcomes, our results add better evidence to prior but not convincing studies that did not find strong evidence for any improvement in nutritional indicators, cognitive abilities or school performance from deworming interventions.[12] The value and impact of deworming programs is an important question to more clearly define, since many health programs include mass drug administration (MDA) for STH infections.

11.1 Internal Validity

Attrition did not bias our sample. To examine whether attrition affected our experimental results, we checked for balance among our "non-missing" cases (those in both the baseline and follow-up surveys). We find that the treatment and control arms are mostly balanced in terms of baseline demographic and household characteristics (Table 2).

The majority of fecal samples from the children were not produced on-site at the time of collection. Thus, it is possible that there was a delay of up to a few hours before children delivered their samples to refrigeration facilities at the school or village clinic. This may have caused an underestimate of total STH prevalence, especially with respect to hookworm; therefore, estimates presented in this paper can be considered a lower bound for actual infection prevalence and infection intensity in our study

population. In addition, this study is focused on examining short-run impacts of deworming; studies are also needed that follow children with no STH infections for longer intervals of time compared to controls with higher (any) levels of infection intensities.

While we do have information on prevalence of STH infection in the sample areas, we do not have information on the intensity of infection in these areas, which could have enlightened us as to whether infected individuals would be likely to have experienced acute symptoms.

The potential for spillovers to threaten internal validity is minimal. Since randomization took place at the township level, within-township spillovers could not affect the control group. We also believe that cross-township spillovers are unlikely given the context. Our sample townships were located in a rural, mountainous region of Guizhou province where the majority of roads are unpaved and travel and communication between villages is consequently limited. Given limited travel between townships, deworming in treatment townships is unlikely to influence infection risk in neighboring townships. Unlike treatment of water-borne parasites (such as schistosomiasis, which is no longer prevalent in the study area), treatment externalities for soil-transmitted helminths are unlikely to take place across large areas. The most likely avenue through which cross-township spillovers might occur would be if village clinicians by chance happened to convene in a central location (i.e., the county) and shared information about the interventions. To reduce the possibility of this happening, we organized the sample so that there was only one to two villages per township. The chance of two village clinicians from different townships ever encountering one another

is extremely low. We also chose our sample so that no two townships are closer than 10 km apart to minimize the risk of contamination due to externalities of the deworming treatment.

Effects due to evaluation itself are a potential concern for the study given that blinding of study participants was impossible in this context. (This is true of any RCT that cannot be double-blinded.) We do not believe John Henry effects are a significant concern given that the control group was unaware of the study intervention; however, it is possible that participants in the treatment groups (and CDC implementers) changed their behavior due to knowledge of participation in a novel government program. While this may pose some challenge to external validity, the possibility of such effects was unfortunately unavoidable in this context. Should the CDC decide to continue the programs in sample townships after the conclusion of the endline survey, however, it may be possible to test for experimenter effects through a long-run follow-up survey at a later date.

To minimize any potential Hawthorne effects due to measurement, the same survey protocol was followed in all sample townships. The pre-baseline stool sampling, baseline survey, and endline survey took place in all townships, including the control group. The stool sampling and baseline survey took place before treatment assignments have been made so it is not possible for enumerator knowledge of treatment status to influence the behavior of participants during the study. Because the evaluation team made additional visits to the Health Education villages, we took care to visit the Control and Deworming villages as well, at the same time.

11.2 External Validity

There are a few limitations of our study that should be addressed if the intervention were to be scaled up or redone in another environment. For example, the majority of fecal samples from the children were not produced on-site at the time of collection. Thus, it is possible that there was a delay of up to a few hours before children delivered their samples to refrigeration facilities at the school or village clinic. This may have caused an underestimate of total STH prevalence, especially with respect to hookworm; therefore, estimates presented in this paper can be considered a lower bound for actual infection prevalence and infection intensity in our study population.

Additionally, reinfection dynamics often cause levels of STH prevalence to persist in a population, a factor that should be considered when evaluating the impact of future deworming interventions. The pattern of rapid reinfection after antihelminthic treatment (in our case we tested the follow-up fecal samples 6 months after treatment) has been consistently identified in geographically diverse populations in both children and adults, [41] and was previously confirmed in our study area.[8, 18] In our scenario, we surmise that children in the intervention group were cleared of STH infection after each of the two rounds of antihelminthic treatment, but experienced high rates of reinfection within the subsequent months leading up to post-intervention evaluation. In a 2012 article by Jia et al., helminth reinfections occur rapidly after treatment, particularly for *A. lumbricoides* and *T. trichiura*. [42] Hence, there is a need for frequent anthelmintic drug administrations to maximize the benefit of preventive chemotherapy. Integrated control approaches emphasizing health education and environmental sanitation are needed to interrupt transmission of STH. Additionally, natural fluctuations in infection in the environment, as previously observed in other populations, [42, 43, 44] may have caused the decrease in

infection prevalence in the control group and further attenuated the between-group difference in prevalence reduction.

While there was no observed impact on any of the nutrition, cognition or educational outcomes, our results are consistent with the subset of studies in the 2012 Cochrane systematic review that also did not find strong evidence for any improvement in nutritional indicators, cognitive abilities or school performance from deworming interventions.[12] One explanation for our results may be that the impact of deworming was attenuated by the low levels of infection intensity in our sample population. The WHO classifies fecal egg counts of 1 to 4,999 eggs per gram (epg) of feces for *Ascaris* and 1 to 999 epg for *Trichuris* as "light-intensity infections". The mean fecal egg count in our sample population (in which the vast majority of infections were with *Ascaris* and *Trichuris*) at baseline was 493.68 epg in the control group and 702.71 epg in the intervention group, placing our sample population in the low range of light-intensity infections. In our case, it is possible that due to low baseline levels of infection intensity in our sample population, we observed no significant impact of deworming on the nutrition, cognition, and education outcome variables. Varying levels of infection intensity could also, in part, explain the variance in the results of the deworming trials included in the 2012 Cochrane review. Perhaps it is the case that studies that found an impact of deworming on nutrition, cognitive abilities and school performance outcomes were more likely to be conducted in areas of (moderate to) high intensity STH infection, while studies that found no impact on key outcomes were conducted in areas of light intensity infection. Unfortunately, this is only speculation because among the trials included in the 2012 Cochrane systematic review on deworming drugs and their effects

on key outcomes, the majority of studies (36 out of 41) failed to report infection intensity in their sample populations; the few studies that did report intensity had exceptionally small sample sizes, utilized different anti-helminthic treatments (i.e. albendazole, pyrantel, mebendazole), and had a high degree of variability in study design.[12] In summary, then, because most research teams did not report infection intensities, it is impossible to know if the baseline STH intensities are able to explain observed differences in the Cochrane studies.

11.3 Future Research

Our findings reveal the opportunity for future randomized controlled trials to examine whether the effect of deworming is empirically associated with baseline infection intensity in the targeted population. These trials should be conducted in settings with varying baseline levels of STH infection intensity in the population, involve uniform implementation of the intervention, and maintain consistency in the measurement and reporting of a comprehensive range of outcomes for rigorous comparison.

The main implications of our study include the following: (1) future deworming studies should quantify and report infection intensity (fecal egg counts) for accurate epidemiological characterization of the sample population; and (2) evidence from future randomized-controlled trials is needed to assess the effect of deworming on key outcomes in populations with moderate and high-intensity STH infections.

12. Specific findings for policy and practice

Our findings have three major implications that inform us about the value of our deworming intervention, which was implemented by local health practitioners in a policy-mimicking procedure. First, addressing policymakers, we find no empirical evidence that the intervention can be justified on the basis of improvement of nutritional indicators, cognitive abilities or school performance. The question remains: should the Chinese government allocate resources for deworming in areas with light-intensity STH infection? If resources are not severely limited, deworming, which is a relatively inexpensive endeavor, will reduce STH infections and, according to the correlations, may help improve cognitive and educational indicators of children with high intensity STH infections (even though the degree of rise may be too small to detect in a study of the scale of this current study). However, the results of this effectiveness study indicate that the intervention is unlikely to produce a rapid impact on nutritional outcomes. Thus, if resources are scarce, public health efforts might best be concentrated on other interventions—though the final decision ought to be supported by detailed cost-benefit analysis.

Second, addressing key influencers, including future researchers, we emphasize the significance of measuring and reporting fecal egg counts in order to categorize infection intensity in the study population. Most deworming trials do not report infection intensity; however, specification of the level of STH infection intensity in the area allows for more accurate characterization of the sample population, and would provide greater consistency when comparing trial results conducted among different populations and geographical areas.

Finally, we hope that future researchers will understand that while our evidence shows that the intervention had no impact on nutritional indicators, cognitive abilities, or school performance in this lightly-infected population, the impact on populations with moderate-to-high levels of infection intensity is an area for further investigation. Our additional descriptive analysis identified a significant correlation between higher levels of infection intensity and worse measures on key outcomes of cognition and school performance (*Table 5*). Our findings reveal the opportunity for future randomized controlled trials to examine whether the effect of deworming is empirically associated with baseline infection intensity in the targeted population. These trials should be conducted in settings with varying baseline levels of STH infection intensity in the population, involve uniform implementation of the intervention, and maintain consistency in the measurement and reporting of a comprehensive range of outcomes for rigorous comparison.

To summarize, the randomized-controlled trial conducted in rural Guizhou, China found that in a population of schoolchildren with light-intensity *Ascaris, Trichuris*, and hookworm infection, a biannual deworming intervention reduced STH infection prevalence and intensity in the population, but had no impact on outcomes of nutrition, cognitive abilities, nor school performance. The results of this effectiveness trial are relevant to developing an effective strategy to reduce soil-transmitted helminth infection and improve the health of children in China and other countries with high STH prevalence. The main implications of our study include the following: (1) in areas with light-intensity STH infection, limited resources might best be concentrated on targeting other, more impactful, public health issues; (2) future deworming studies should quantify

and report infection intensity (fecal egg counts) for accurate epidemiological characterization of the sample population; and (3) evidence from future randomized-controlled trials is needed to assess the effect of deworming on key outcomes in populations with moderate and high-intensity STH infections.

Thus, the results of this study suggest that a comprehensive deworming program to reduce STH infection rates in rural China should include the following components: free deworming treatment that is provided annually either through schools or village clinics, educational materials that provide accurate and necessary knowledge about STH, an emphasis on the impacts of deworming on children's educational attainment and future financial prospects, education about STH infections through community engagement, cultural sensitivity when overturning local myths about STH infections and deworming treatment, and lastly, greater efforts towards improving the quality of the village health care system.

Appendix: Figures and Tables

Figure 1: Covers of STH educational pamphlets

Figure 2: Timeline of the project.

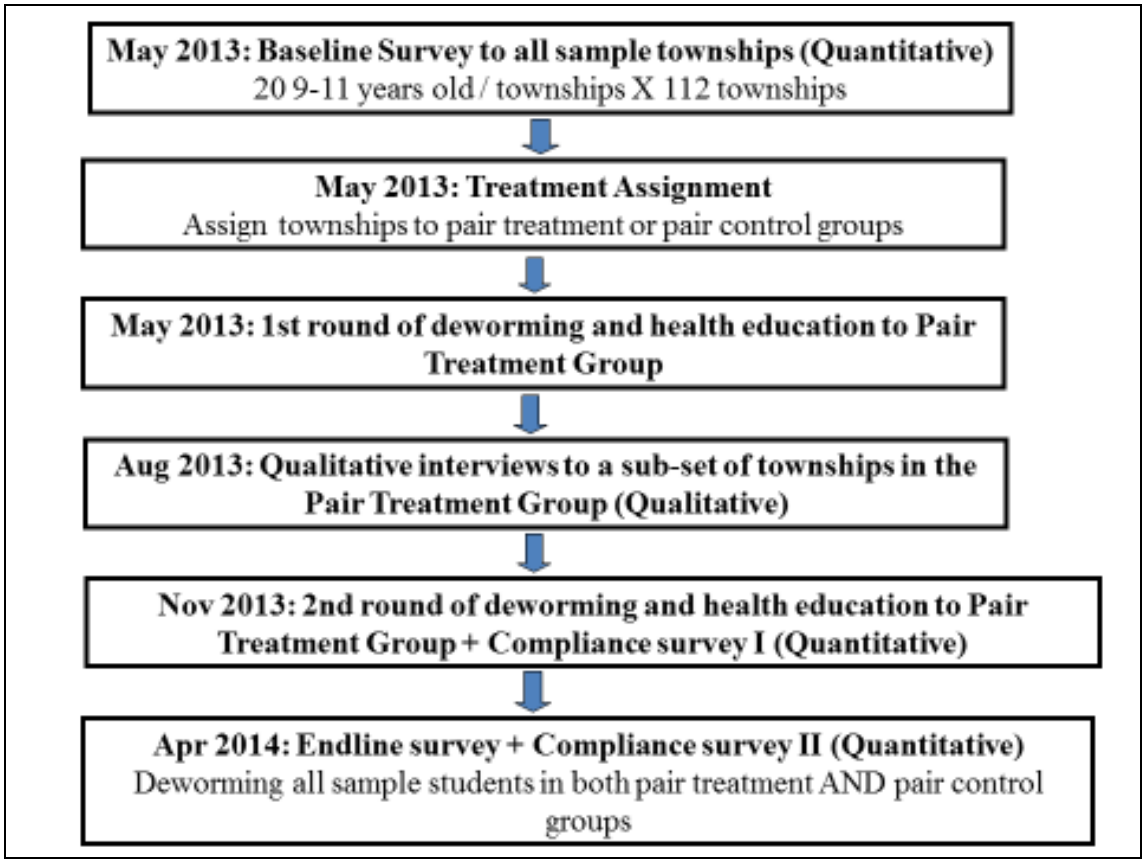

Figure 3: Trial profile

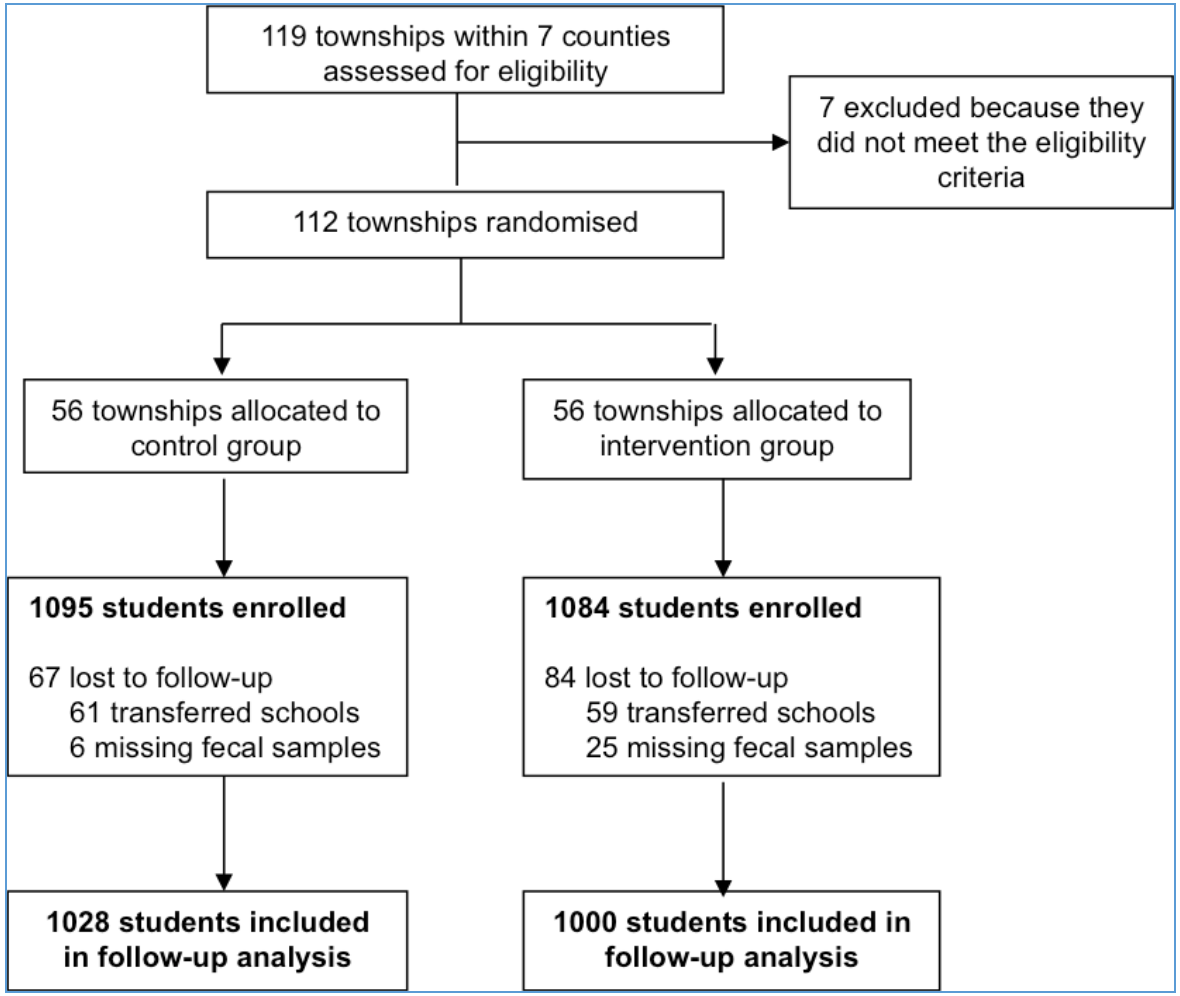

Table 1: WISC-IV Core Tests of Cognitive Ability

Test	Method	Type	Skills
Digit span	Proctor verbally states numbers; child repeats them back in same order and in inverse order.	Verbal test	Measures auditory short-term memory, sequencing skills, attention, and concentration
Letter Number Sequencing	Proctor verbally states sequences of random letters and digits; child repeats the digits in numerical order then the letters in alphabetical order	Verbal test	Measures sequencing, mental manipulation, attention, short-term auditory memory
Coding	Child is shown a legend where numbers of signs are associated with shapes; child is presented with scenarios involving matching according to the legend within a specific time limit (120 seconds)	Non-verbal test, performance test	Measures processing speed, short-term memory, perceptual abilities, motor coordination, speed
Symbol search	Child scans a search group and indicates whether the target symbol(s) matches any of the symbols in the search group within a specific time limit (120 seconds)	Non-verbal test, performance test	Measures processing speed, short-term visual memory, visual-motor coordination

Source: WISC-IV, Wechsler Intelligence Scale for Children-Fourth Edition.[8]

Table 2: Baseline demographic and household characteristics of study participants

	Control (N = 1095), 95% CI[1]	Intervention (N = 1084), 95% CI	P value
Individual characteristics			
Age	10.61 (10.56 to 10.66)	10.56 (10.50 to 10.61)	0.391
Female (%)	43.38 (40.44 to 46.32)	48.99 (46.00 to 51.97)	**0.044***
Boarding at school (%)	27.38 (24.74 to 30.02)	24.84 (22.26 to 27.41)	0.680
Dong ethnic minority (%)	47.03 (44.07 to 49.99)	43.54 (40.59 to 46.50)	0.675
Miao ethnic minority (%)	36.07 (33.22 to 38.92)	37.73 (34.84 to 40.62)	0.830
Shui ethnic minority (%)	2.92 (1.92 to 3.92)	4.43 (3.20 to 5.65)	0.597
Household characteristics			
Number of siblings	1.13 (1.07 to 1.18)	1.24 (1.18 to 1.29)	0.129
Pieces of durable assets	8.45 (8.27 to 8.63)	8.30 (8.11 to 8.50)	0.581
Parents are migrant workers (%)	31.6 (28.85 to 34.34)	28.53 (25.84 to 31.22)	0.340
Mother attended secondary school (%)	7.28 (5.75 to 8.82)	6.74 (5.25 to 8.24)	0.690
Father attended secondary school (%)	12.26 (10.32 to 14.19)	10.99 (9.13 to 12.85)	0.428
Sanitation and hygiene			
Washes hands before eating (%)	84.63 (82.49 to 86.77)	84.24 (82.08 to 86.41)	0.854
Washes hands after using toilet (%)	87.75 (85.81 to 89.68)	85.74 (83.66 to 87.83)	0.318
Drinks boiled water only (%)	5.21 (3.89 to 6.52)	8.39 (6.74 to 10.05)	**0.013***
Wears shoes while playing outside (%)	32.33 (29.55 to 35.10)	33.30 (30.49 to 36.11)	0.771
House has dirt floor (%)	17.08 (14.85 to 19.31)	14.30 (12.21 to 16.39)	0.303
House has dirt-based latrine (%)	19.58 (17.23 to 21.93)	23.38 (20.86 to 25.90)	0.160
Family uses feces as fertilizer (%)	65.48 (62.66 to 68.30)	62.08 (59.19 to 64.98)	0.270
Infection Prevalence			
Any STH infection (%)	41.10 (38.18 to 44.01)	42.62 (39.67 to 45.57)	0.779
Ascaris infection (%)	30.50 (27.77 to 33.23)	31.09 (28.33 to 33.85)	0.891
Trichuris infection (%)	23.29 (20.78 to 25.80)	24.35 (21.80 to 26.91)	0.847
Hookworm infection (%)	1.00 (0.41 to 1.60)	0.74 (0.23 to 1.25)	0.568
Ascaris and *Trichuris* co-infection (%)	12.97 (10.98 to 14.96)	12.92 (10.92 to 14.91)	0.989
Infection Intensity (among samples with positive infection)			
Ascaris infection (epg)	728.32 (526.90 to	1065.04 (741.17 to 1388.92)	0.151

	929.75)		
Trichuris infection (epg)	55.90 (38.29 to 73.50)	71.79 (48.01 to 95.58)	0.562
Hookworm infection (epg)	17.33 (-16.83 to 51.50)	18.00 (-70.94 to 106.94)	0.967

* **Bolded** values indicate significance at 95% CI.
[1] CI denotes confidence interval. P-values adjusted for clustering at the township level.

Table 3: Infection prevalence and intensity in control and intervention groups

Variable	**Control Group Mean, sd (CI)**	**Intervention Group**	**Intervention Effect (95% CI[1]),**			
			Unadjusted, RMSE	**P value**	**Adjusted[2], RSME**	**P value**
Infection Prevalence (%)						
Baseline	41.10, 1.49 (38.18, 44.01)	42.62, 1.50 (39.67, 45.57)	1.06 (0.69 to 1.65)	0.779	1.15 (0.93 to 1.43)	0.192
Follow-up	31.40, 1.45 (28.56 to 34.23)	27.66, 1.41 (24.89 to 30.43)	0.84 (0.52 to 1.35)	0.464	0.71 (0.52 to 0.96)	**0.026***
Infection Intensity[3] (epg)						
Baseline	493.68, 68.96 (357.83 to 629.53)	702.71, 107.45 (491.07 to 914.34)	209.02 (-167.29 to 585.33), 1420.4	0.272	115.29 (-101.48 to 332.06), 1352.4	0.293
Follow-up	533.32, 72.46 (390.69 to 675.95)	299.93, 53.78 (193.98 to 405.89)	-233.39 (-489.36 to 22.58), 1065.2	0.073	-209.78 (-383.16 to -36.39), 766.02	**0.018***

* **Bolded** values indicate significance at 95% CI.

[1] CI denotes confidence interval.

[2] Values were adjusted for individual characteristics (gender, age, boarding status, minority identification) and household characteristics (siblings, durable assets, parental migrant worker status, parental education levels), as well as township pair-fixed effects. Coefficients for infection prevalence are reported as an odds-ratio.as well as township pair-fixed effects. In the case of follow-up, values were also adjusted for the baseline value of the dependent variable. Coefficients for infection prevalence are reported as odds-ratio.

[3] Infection intensity calculated as average fecal egg count among samples with positive infection.

Table 4: Differences in outcomes of nutrition, cognitive abilities, and school performance between control and intervention groups

Variable	**Control Group** **Mean, sd (CI)**	**Intervention Group**	**Intervention Effect (95% CI)**			
			Unadjusted, RMSE	**P value**	**Adjusted[2], RSME**	**P value**
Nutritional Indicators						
Hemoglobin levels						
Baseline	126.17, 0.38 (125.43 to 126.91)	126.33, 0.38 (125.59 to 127.07)	0.16 (-1.73 to 2.05), 12.449	0.864	0.24 (-0.67 to 1.14), 11.722	0.604
Follow-up	132.16, 0.40 (131.37 to 132.95)	131.73, 0.41 (130.92 to 132.54)	-0.43 (-2.53 to 1.68), 12.985	0.690	-0.33 (-1.64 to 0.98), 11.193	0.623
Anemia prevalence (%)						
Baseline	16.62, 1.13 (14.41 to 18.83)	16.14, 1.11 (13.95 to 18.34)	0.97 (0.70 to 1.34)	0.833	0.93 (0.79 to 1.10)	0.385
Follow-up	9.98, 0.93 (8.14 to 11.82)	11.58, 1.00 (9.61 to 13.55)	1.18 (0.79 to 1.76)	0.413	1.25 (0.91 to 1.72)	0.174
% Stunted (HAZ < -2)						
Baseline	26.98, 1.36 (24.32 to 29.65)	29.66, 1.41 (26.89 to 32.42)	1.14 (0.86 to1.52)	0.368	1.10 (0.92 to 1.31)	0.291
Follow-up	23.48, 1.34 (20.85 to 26.11)	27.63, 1.42 (24.84 to 30.42)	1.24 (0.93 to 1.67)	0.148	1.15 (0.85 to 1.55)	0.367
% Underweight (WAZ < -2)						
Baseline	24.11, 1.29 (21.57 to 26.65)	28.90, 1.38 (26.20 to 31.61)	1.28 (1.01 to 1.63)	**0.045***	1.29 (1.09 to 1.54)	**0.004***
Follow-up	21.37, 1.28 (18.85 to 23.88)	24.19, 1.34 (21.56 to 26.82)	1.17 (0.92 to 1.51)	0.204	0.77 (0.56 to 1.06)	0.113
Cognitive Abilities						

Processing Speed Index Score						
Baseline	86.21, 0.40 (85.44 to 86.99)	86.09, 0.40 (85.31 to 86.87)	-0.12 (-2.47 to 2.23), 13.11	0.919	0.16 (-1.28 to 1.59), 12.166	0.827
Follow-up	88.18, 0.41 (87.37 to 88.99)	88.83, 0.42 (88.01 to 89.65)	0.65 (-1.47 to 2.77), 13.281	0.545	0.63 (-0.22 to 1.49), 10.157	0.143
Working Memory Index Score						
Baseline	78.68, 0.30 (78.08 to 79.27)	78.51, 0.30 (77.92 to 79.10)	-0.16 (-1.60 to 1.28), 9.9338	0.822	-0.05 (-0.98 to 0.89), 9.5037	0.922
Follow-up	78.23, 0.32 (77.61 to 78.86)	78.50, 0.33 (77.86 to 79.14)	0.27 (-1.18 to 1.72), 10.309	0.715	0.51 (-0.09 to 1.11), 8.1718	0.093
School Performance						
School attendance rate (%)						
Baseline	86.73, 1.06 (84.65 to 88.81)	87.32, 1.04 (85.29 to 89.35)	1.05 (0.68 to 1.64)	0.818	1.08 (0.75 to 1.56)	0.692
Follow-up	86.13, 1.18 (83.81 to 88.45)	85.30, 1.20 (82.95 to 87.66)	0.93 (0.57 to 1.54)	0.790	0.86 (0.55 to 1.33)	0.496
Normalized TIMSS score						
Baseline	0.00, 0.03 (-0.06 to 0.06)	-0.04, 0.03 (-0.10 to 0.02)	-0.04 (-0.22 to 0.14), 0.99424	0.633	0.01 (-0.1 to 0.11), 0.85899	0.912
Follow-up	0.00, 0.03 (-0.06 to 0.06)	-0.07, 0.03 (-0.14 to -0.01)	-0.07 (-0.24 to 0.10), 1.0339	0.412	-0.04 (-0.09 to 0.02), 0.64296	0.190

* **Bolded** values indicate significance at 95% CI.

[1] CI denotes confidence interval.

[2] Values were adjusted for individual characteristics (gender, age, boarding status, minority identification) and household characteristics (siblings, durable assets, parental migrant worker status, parental education levels), as well as township pair-fixed effects. In the case of follow-up, values were also adjusted for the baseline value of the dependent variable. Coefficients for anemia prevalence, % stunted, % underweight and school attendance rate are reported as an odds-ratio.

Table 5: Correlation between infection intensity (fecal egg counts) and outcomes at baseline among samples with positive infection

Outcome Variables	Correlation Coefficient
% Anemic	0.0125 (0.7823)
% Stunted	0.0844 (0.0640)
% Underweight	0.0862 (0.0566)
Processing Speed Index	**-0.1217* (0.0070)**
Working Memory Index	**-0.1288* (0.0043)**
School Attendance	-0.0279 (0.5452)
Normalized TIMSS Score	**-0.1416* (0.0017)**

* **Bolded** values indicate significance at 95% CI.
[1] Brackets contain p-values

Table 6: Average-treatment-effect-on-treated[a] (ATT) analysis of differences in infection, outcomes of nutrition, cognitive abilities, and school performance between control and intervention groups

Variable	Intervention Effect (95% CI[b])			
	Unadjusted	**P value**	**Adjusted[c]**	**P value**
Infection Characteristics				
Infection Prevalence (%)				
Baseline	0.07 (-0.45 to 0.60)	0.779	0.15 (-0.09 to 0.39)	0.209
Follow-up	-0.21 (-0.73 to 0.31)	0.435	-0.39 (-0.69 to -0.09)	**0.011***
Infection Intensity (epg) (among samples with positive infection)				
Baseline	365.05 (-290.39 to 1020.49)	0.271	191.98 (-171.49 to 555.46)	0.296
Follow-up	-512.58 (-1077.22 to 52.05)	0.075	-370.37 (-679.12 to -61.62)	**0.019***
Nutritional Indicators				
Hemoglobin levels				
Baseline	0.31 (-3.31 to 3.93)	0.865	0.45 (-1.26 to 2.15)	0.604
Follow-up	-0.77 (-4.57 to 3.03)	0.689	-0.58 (-2.93 to 1.76)	0.622
Anemia prevalence (%)				
Baseline	-0.04 (-0.38 to 0.30)	0.819	-0.07 (-0.25 to 0.11)	0.443
Follow-up	0.15 (-0.22 to 0.52)	0.434	0.20 (-0.10 to 0.50)	0.185
% Stunted (HAZ < -2)				
Baseline	0.15 (-0.18 to 0.48)	0.370	0.10 (-0.10 to 0.30)	0.339
Follow-up	0.23 (-0.09 to 0.55)	0.154	0.14 (-0.14 to 0.42)	0.335
% Underweight (WAZ < -2)				
Baseline	0.28 (0.00 to 0.56)	0.050	0.28 (0.08 to 0.48)	**0.007***
Follow-up	0.17 (-0.09 to 0.43)	0.209	-0.21 (-0.50 to 0.09)	0.174
Cognitive Abilities				
Processing Speed Index Score				
Baseline	-0.23 (-4.74 to 4.28)	0.919	0.30 (-2.41 to 3.01)	0.827
Follow-up	1.17 (-2.66 to 5.01)	0.545	1.14 (-0.39 to 2.67)	0.142

Working Memory Index Score				
Baseline	-0.31 (-3.07 to 2.44)	0.822	-0.09 (-1.86 to 1.68)	0.922
Follow-up	0.49 (-2.15 to 3.12)	0.715	0.92 (-0.15 to 1.99)	0.093
School Performance				
School attendance rate (%)				
Baseline	0.05 (-0.4 to 0.51)	0.818	0.06 (-0.3 to 0.43)	0.736
Follow-up	-0.07 (-0.57 to 0.44)	0.796	-0.15 (-0.59 to 0.28)	0.491
Normalized TIMSS score				
Baseline	-0.08 (-0.43 to 0.26)	0.632	0.01 (-0.19 to 0.21)	0.912
Follow-up	-0.13 (-0.43 to 0.18)	0.412	-0.07 (-0.17 to 0.03)	0.187

* **Bolded** values indicate significance at 95% CI.

[a] Average-treatment-effect-on-treated (ATT) were estimated by using the initial random treatment assignment as the instrument variable for the observed compliance (whether a student took the 2 albendazole pills passed by the intervention team of the project.

[b] CI denotes confidence interval.

[c] Values were adjusted for individual characteristics (gender, age, boarding status, minority identification) and household characteristics (siblings, durable assets, parental migrant worker status, parental education levels), as well as township pair-fixed effects. In the case of follow-up, values were also adjusted for the baseline value of the dependent variable.

Appendix Table 1. Power Calculations

Outcome	**Within-township Sample**	**ICC Assumption (ρ)**	**R2 Assumption (R^2)**	**Effect Size (δ)**	**Power at 110 townships per comparison**	**Required Sample per comparison for 80% Power**
Educational Outcomes						
TIMMS	20 4th and 5th graders (9-11 year olds)	0.16	0.5	0.25	96%	63 townships
TIMMS	20 4th and 5th graders (9-11 year olds)	0.16	0.5	0.20	84%	100 townships
Health Outcomes						
Worm Burden	20 4th and 5th graders	0.07	0.5	0.25	100%	36 townships
Worm Burden	20 4th and 5th graders	0.07	0.5	0.20	98%	55 townships
Hemoglobin Concentration	20 4th and 5th graders	0.16	0.5	0.25	97%	56 townships
Hemoglobin Concentration	20 4th and 5th graders	0.16	0.5	0.20	88%	89 townships
Weight-for-age Z-score	20 4th and 5th graders	0.01	0.6	0.20	100%	32 townships
Weight-for-age Z-	20 4th and 5th	0.01	0.6	0.15	98%	54 townships

score	graders					
Height-for-age Z-score	20 4th and 5th graders	0.05	0.6	0.20	99%	44 townships
Height-for-age Z-score	20 4th and 5th graders	0.05	0.6	0.15	92%	74 townships
BMI-for-age Z-score	20 4th and 5th graders	0.03	0.6	0.20	100%	37 townships
BMI-for-age Z-score	20 4th and 5th graders	0.03	0.6	0.15	96%	66 townships
WISC Test	20 4th and 5th graders	0.12	0.4	0.25	97%	62 townships
WISC Test	20 4th and 5th graders	0.12	0.4	0.20	86%	94 townships

Appendix Table 2. Power Calculations for Sub-group Analyses

Outcome	**Within-township Sample**	**ICC Assumption (ρ)**	**R2 Assumption (R2)**	**Effect Size (δ)**	**Power at 110 townships per comparison**	**Required Sample per comparison for 80% Power**
Worm Burden	20 4th and 5th graders	0.06	0.5	0.25	99%	41 townships
Worm Burden	20 4th and 5th graders	0.06	0.5	0.20	96%	63 townships
Hemoglobin Concentration	20 4th and 5th graders	0.16	0.5	0.25	96%	62 townships
Hemoglobin Concentration	20 4th and 5th graders	0.16	0.5	0.20	85%	100 townships
Weight-for-age Z-score	20 4th and 5th graders	0.01	0.6	0.20	99%	45 townships
Weight-for-age Z-score	20 4th and 5th graders	0.01	0.6	0.15	92%	78 townships
Height-for-age Z-score	20 4th and 5th graders	0.07	0.6	0.20	97%	61 townships
Height-for-age Z-score	20 4th and 5th graders	0.07	0.6	0.15	81%	106 townships
BMI-for-age Z-score	20 4th and 5th graders	0.01	0.6	0.20	99%	45 townships
BMI-for-age Z-score	20 4th and 5th graders	0.01	0.6	0.15	92%	76 townships

Appendix Table 3: Effects of Intervention on Outcome Variables, by worm load

Outcome variable in the follow-up	**Intervention Effect based on adjusted model**			
	Sub-sample with worm count less than 1000 epg (n=408)		**Sub-sample with worm count no less than 1000 epg (n=82)**	
	Coefficient[1] (95% CI[2])	**p-value**	**Coefficient (95% CI)**	**p-value**
Infection				
Infection Intensity[3] (epg)	-195.75 (-295.04, -96.45)	**0.000***	-561.51 (-1370.82, 247.81)	0.163
Nutritional indicators				
Hemoglobin levels	-3.65 (-6.186, -1.11)	**0.005***	-2.15 (-11.33, 7.03)	0.636
Anemia prevalence (%)	3.17 (1.53, 6.57)	**0.002***	NA	
% Stunted (HAZ < -2)	0.99 (0.48, 2.01)	0.968	0.03 (0.00, 0.82)	**0.038***
% Underweight (WAZ < -2)	0.23 (0.08, 0.66)	**0.006***	NA	
Cognitive Abilities				
Processing Speed Index Score	-1.07 (-3.28, 1.15)	0.340	-0.70 (-9.07, 7.67)	0.865
Working Memory Index Score	0.52 (-0.96, 2.00)	0.489	3.65 (0.04, 7.26)	**0.048***
School Performance				
School attendance rate (%)	0.73 (0.37, 1.41)	0.348	13.36 (0.02, 7745.96)	0.425
Normalized TIMSS score	-0.19 (-0.31, -0.07)	**0.002***	-0.15 (-0.50, 0.21)	0.403

* **Bolded** values indicate significance at 95% CI.

[1] Values were adjusted for individual characteristics (gender, age, boarding status, minority identification) and household characteristics (siblings, durable assets, parental migrant worker status, parental education levels), as well as township pair-fixed effects. Coefficients for infection prevalence are reported as an odds-ratio.as well as township pair-fixed effects. In the case of follow-up, values were also adjusted for the baseline value of the dependent variable. Coefficients for infection prevalence are reported as odds-ratio.

[2] CI denotes confidence interval.

[3] Infection intensity calculated as average fecal egg count among samples with positive infection.

NA stands for not available due to small sample size and missing information.

Appendix Table 4: Effects of Intervention on Outcome Variables, by migration status of parents

Outcome variable in the follow-up	**Intervention Effect based on adjusted model**			
	Treated		**Interaction of treated X boarder**	
	Coefficient[1] (95% CI[2])	**p-value**	**Coefficient (95% CI)**	**p-value**
Infection				
Infection Prevalence (%)	0.58 (0.40, 0.82)	**0.002***	2.12 (0.99, 4.54)	0.052
Infection Intensity[3] (epg)	-256.50 (-487.95, -25.05)	**0.030***	176.36 (-163.85, 516.57)	0.306
Nutritional indicators				
Hemoglobin levels	-0.629 (-2.21, 0.95)	0.432	1.18 (-1.50, 3.86)	0.384
Anemia prevalence (%)	1.37 (0.90, 2.08)	0.141	0.66 (0.29, 1.52)	0.334
% Stunted (HAZ < -2)	1.13 (0.78, 1.64)	0.531	1.07 (0.49, 2.35)	0.862
% Underweight (WAZ < -2)	0.72 (0.50, 1.04)	0.080	1.27 (0.57, 2.87)	0.560
Cognitive Abilities				
Processing Speed Index Score	0.85 (-0.18, 1.88)	0.105	-0.85 (-3.33, 1.62)	0.497
Working Memory Index Score	0.87 (0.20, 1.55)	**0.012***	-1.42 (-3.22, 0.37)	0.119
School Performance				
School attendance rate (%)	1.07 (0.65, 1.76)	0.800	0.42 (0.18, 0.98)	**0.045***
Normalized TIMSS score	0.01 (-0.07, 0.078)	0.869	-0.18 (-0.31, -0.04)	**0.010***

* **Bolded** values indicate significance at 95% CI.

[1] Values were adjusted for individual characteristics (gender, age, boarding status, minority identification) and household characteristics (siblings, durable assets, parental migrant worker status, interaction term between treatment and parental migration status, parental education levels), as well as township pair-fixed effects. Coefficients for infection prevalence are reported as an odds-ratio.as well as township pair-fixed effects. In the case of follow-up, values were also adjusted for the baseline value of the dependent variable. Coefficients for infection prevalence are reported as odds-ratio.

[2] CI denotes confidence interval.

[3] Infection intensity calculated as average fecal egg count among samples with positive infection.

Appendix Table 5: Crosstab of the self-reported number of deworming pills taken by the treatment group students, number of students

	# of pills taken in Nov treatment			
	0	1	2	Sub-total
# of pills taken in May treatment				
0	89	31	36	156
1	27	63	85	175
2	78	109	566	753
subtotal	194	203	687	1084

References

1. Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ, et al. Helminth infections: the great neglected tropical diseases. J Clin Invest 2008; **118**: 1311–1321.
2. Pullan RL, Brooker SJ. The global limits and population at risk of soil-transmitted helminth infections in 2010 Parasit. Vectors 2012; **5**: 81.
3. Utzinger J, Becker SL, Knopp S, Blum J, Neumayr AL, Keiser J, Hatz CF. Neglected tropical diseases: diagnosis, clinical management, treatment and control Swiss Med. Wkly. 2012; **142**:13727.
4. Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, et al. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. Lancet 2006; **367**:1521–1532.
5. Hall A, Hewitt G, Tuffrey V. A review and meta-analysis of the impact of intestinal worms on child growth and nutrition. Matern. Child Nutr. 2008; **4**:118–236.
6. Suchdev PS, Davis SM, Bartoces M, et al. Soil-transmitted helminth infection and nutritional status among urban slum children in Kenya. Am J Trop Med Hyg 2014; **90**:299-305.
7. Ezeamama AE, Friedman JF, Acosta LP, Bellinger DC, Langdon GC, Manalo DL, Olveda RM, Kurtis JD, McGarvey ST. Helminth infection and cognitive impairment among Filipino children. Am J Trop Med Hyg 2005; **72**:540–548.
8. Liu C, Luo R, Yi H, Zhang L, Li S, Bai Y, Medina A, Rozelle S, Smith S, Wang G, Wang J. Soil-transmitted helminths in southwestern China: a cross-sectional study of links to cognitive ability, nutrition, and school performance among children. PLoS Negl Trop Dis 2015; **9**: e0003877. doi:10.1371/journal.pntd.0003877.
9. Miguel E, Kremer M. Worms: Identifying impacts on education and health in the presence of treatment externalities. Econometrica 2004;**72**:159–217.
10. Vercruysse J, Behnke JM, Albonico M, Ame SM, Angebault C, et al. Assessment of the anthelmintic efficacy of albendazole in school children in seven countries where soil-transmitted helminths are endemic. PLoS Negl Trop Dis 2011; **5**: e948.
11. Keiser J, Utzinger J. Efficacy of current drugs against soil-transmitted helminth infections. J Amer Med Assoc 2008; **299**:1937–1948.
12. 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 (Review). The Cochrane Library 2012; 11. The Cochrane Collaboration.
13. Ebenezer R, Gunawardena K, Kumarendran B, Pathmeswaran A, Jukes MCH, Drake LJ, de Silva N. Cluster-randomised trial of the impact of school-based

deworming and iron supplementation on the cognitive abilities of schoolchildren in Sri Lanka's plantation sector. Tropical Medicine and International Health 2013; **8**: 942–951.

14. Yap P, Wu F-W, Du Z-W, Hattendorf J, Chen R, et al. Effect of deworming on physical fitness of school-aged children in Yunnan, China: a double- blind, randomized, placebo-controlled trial. PLoS Negl Trop Dis 2014; **8**(7): e2983. doi:10.1371/journal.pntd.0002983
15. 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.
16. Vicotra CG, Black RE, Boerma JT, Bryce J. Measuring impact in the Millennium Development Goal era and beyond: a new approach to large-scale effectiveness evaluations. Lancet 2011; **377**:85–95.
17. Gartlehner G, Hansen RA, Nissman D, et al. Criteria for distinguishing effectiveness from efficacy trials in systematic reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006 Apr.
18. Wang X, Zhang L, Luo R, Wang G, Chen Y, et al. Soil-transmitted helminth infections and correlated risk factors in preschool and school-aged children in rural southwest China. PLoS ONE 2012; **7**(9): e45939. doi:10.1371/journal.pone.0045939
19. Zhou, H, Ohtsuka R, He Y, Yuan L. et al. Impact of parasitic infections and dietary intake on child growth in the schistosomiasis-endemic Dongting Lake region, China. The American Journal of Tropical Medicine and Hygiene, 2005; **72**:534–539.
20. World Health Organization. Irion Deficiency Anemia: Assessment, Prevention and Control, A Guide for Programme Managers. Geneva: World Health Organization. 2001.
21. World Health Organization. WHO AnthroPlus for Personal Computers Manual: Software for Assessing Growth of the World's Children and Adolescents Retrieved from http://www.who.int/growthref/tools/en/. 2009.
22. Centers for Disease Control and Prevention. A SAS Program for the 2000 CDC Growth Charts (ages 0 to <20 years). Retrieved from http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm. 2014.
23. Bruhn M, McKenzie D, In pursuit of balance: randomization in practice in development field experiments. American Economic Journal: Applied Economics 2009; **1**:200–232. doi:10.1257/app.1.4.200
24. Greevy R, Lu B, Silber JH, Rosenbaum P. Optimal multivariate matching before randomization. Biostat 2004; **5**:263–275. doi:10.1093/biostatistics/5.2.263
25. Shang Y. Burden of diseases on soil-transmitted helminth infections among school-age children in China [in Chinese]. Doctoral Thesis, National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention.

Shanghai, China. 2011.

26. Angrist JD, Imbens GW, Rubin D B. Identification of causal effects using instrumental variables. Journal of the American Statistical Association 1996; **91**:444-455.

27. China National Bureau of Statistics 2013. China Statistical Yearbook 2012. China Statistics Press. Beijing, China. 2013.

28. Jamison DT, Breman JG, Measham AR, et al. Disease control priorities in developing countries. 2nd edition. Chapter 24-Helminth infections: soil-transmitted helminth infections and schistosomiasis. World Bank. 2006.

29. Imai K, Gary K, Clayton N. The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation. Statistical Science, 2009; 24: 29–53.

30. Bruhn, Miriam, and David McKenzie. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. American Economic Journal: Applied Economics 1, 2009; 4: 200–232.

31. Guizhou Provincial Bureau of Statistics. Guizhou Statistical Yearbook 2011. Beijing, China: China Statistics Press. 2012.

32. Moore, Ryan T. Block tools: blocking, assignment, and diagnosing interference in randomized experiments. Technical Report. 2012.

33. Liu C, Luo R, Li S, Bai Y, Zhang L, Medina A, Smith DS, Rozelle S. (2014). Soil-Transmitted Helminths in Southwestern China: Links to Cognitive Ability, Nutrition, and School Performance among Children. Rural Education Action Program, working paper.

34. Ash LR, Orihel TC, Savioli L, Sin MA, Montresor A, Renganathan E. Training Manual on Diagnosis of Intestinal Parasites. World Health Organization, Geneva. 1998.

35. World Health Organization. Helminth Control for School-age Children. 2011.

36. de Onis M, Blossner M, Borghi E, Morris R, Fronglio EA. Methodology for estimating regional and global trends in child malnutrition. International Journal Epidemiology 2004; **33**:1260-70.

37. Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV) Administration and Scoring Manual (Chinese version). King-may Psychological Assessment LTD, Zhuhai, China. 2008.

38. Nokes C, Grantham-McGregor SM, Sawyer AW, Cooper ES, Bundy DA. Parasitic helminthic infection and cognitive function in school children. Proc. R. Soc. Lond. B. Biol. Sci. 1992; **247**:77–81.

39. Jardim-Botelho A, Raff S, De Ávila Rodrigues R, Hoffman HJ, Diemert DJ, et al. Hookworm, ascaris lumbricoides infection and polyparasitism associated with poor cognitive performance in Brazilian schoolchildren. Tropical Medicine and International Health 2008; **13**(8):994-1004.

40. Mullis IV, Martin MO, Gonzalez EJ, Chrostowski SJ. TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. International Association for the Evaluation of Educational Achievement, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College; U.S.A., MA: 2004.

41. Holland CV. Predisposition to ascariasis: patterns, mechanisms and implications. Parasitology 2009; **136**(12):1537-47. doi: 10.1017/ S0031182009005952.

42. Jia T-W, Melville S, Utzinger J, King CH, Zhou X-N. Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis. PLoS Negl Trop Dis 2012; **6**(5): e1621. doi:10.1371/journal.pntd.0001621

43. Cabrera BD. Reinfection and infection rates of ascariasis in relation to seasonal variation in the Philippines. Southeast Asian J Trop Med Public Health 1984; **15**:394–401.

44. Pan CT, Ritchie LS, Hunter GW 3rd. Reinfection and seasonal fluctuations of Ascaris lumbricoides among a group of children in an area where night soil is used. J Parasitol 1954; **40**:603–608.

45. World Health Organization. Helminth Control in School-age Children: A Guide for Managers of Control Programmes. Second Edition. 2011.

46. McKinley, T. The Distribution of Wealth in Rural China. Armonk, NY: M.E. Sharpe. Print.1996

47. The State Council Office of China. Guowuyuan Bangongting Guanyu Jinyibu Jiaqiang Xiangchun Yisheng Duiwu Jianshe de Zhidao Yijian, Guobanfa (2011) 31 Hao. 2011[Guidelines Issued by the State Council Office about Further Improving the Capacity of Rural Village Doctors, Guobanfa (2011) No. 31].

48. 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 (Review). The Cochrane Library 2015; 7. The Cochrane Collaboration.

49. Lu L, Liu CF, Zhang LZ, Medina A, Smith S, Rozelle S. Gut Instincts: Knowledge, Attitudes, and Practices Regarding Soil-Transmitted Helminths in Rural China. PLOS Neglected Tropical Diseases. DOI: 10.1371/journal.pntd.0003643.

50. Bieri, Franziska A., Darren J. Gray, Gail M. Williams, Giovanna Raso, Yue-Sheng Li, Liping Yuan, Yongkang He, Robert S. Li, Feng-Ying Guo, Sheng-Ming Li, and Donald P. McManus. 2013, Health-Education Package to Prevent Worm Infections in Chinese Schoolchildren. N Engl J Med 2013; 368:1603-1612, April 25, 2013, DOI: 10.1056/NEJMoa1204885.

51. Ozier, Owen. Exploiting Externalities to Estimate the Long-Run Effect of Early Childhood Deworming, Policy Research Working Paper 7052. October 2014. The World Bank.

52. Baird, Sarah, Joan Hicks, Michael Kremer, Edward Miguel. Worms at work longrun impacts of child health gains. 2011. NBER Working Paper 21428.

53. Aiken, AM, Calum Davey, James R Hargreaves, Richard J Hayes. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: A pure replication. in International Journal of Epidemiology 44(5). July 2015. DOI: 10.1093/ije/dyv127.

54. Davey, Calum, Aiken, AM, Richard J Hayes, James R Hargreaves. 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. International Journal of Epidemiology 44(5). July 2015. DOI: 10.1093/ije/dyv128.