Estimating deworming school participation impacts and externalities in Kenya: A Comment on Aiken *et al.* (2014)[†]

Joan Hamory Hicks, University of California, Berkeley CEGA
Michael Kremer, Harvard University and NBER
Edward Miguel[*], University of California, Berkeley and NBER

October 2014

Abstract: Aiken et al. (2014) usefully correct some errors in Miguel and Kremer (2004). Miguel and Kremer (2004) made two key claims: 1) deworming creates positive epidemiological externalities, thus causing estimates of the impact of deworming based on individual randomization to be biased downwards; and 2) deworming increases school participation. The results in Aiken et al. (2014) are consistent with these findings. In addition to direct impacts of deworming treatment on worm infections, Aiken et al. (2014) find externality effects within schools on untreated pupils, as well as externality effects across schools up to 3 km away. Similarly, with regard to school participation, both Miguel and Kremer (2004) and Aiken et al. (2014) find direct effects of deworming and externality effects within schools on untreated pupils. As Aiken et al. (2014) point out, most of the errors they identify (rounding errors or updates to the data set) lead to only small changes in estimated coefficients. The key change is that Miguel and Kremer (2004) measured externalities among schools located within 3-6 km that were among the 12 closest schools, rather than among all schools within 3-6 km, as reported; Aiken et al. (2014) report results including the full set of schools within 3-6 km. With the updated data, there is no evidence that worm infection externalities extend beyond the 12 closest schools to the full set of schools within 6 km, perhaps unsurprisingly given the local nature of disease transmission. We disagree with Aiken et al.'s (2014) conclusion that "there was no evidence of a between-school indirect effect" or an overall effect of deworming on school participation. We show that this conclusion is based on an approach that adds substantial noise to the estimation, by heavily weighting a non-significant 3-6 km externality estimate. This note addresses these and other points, and comments on the current state of deworming evidence.

[†] Acknowledgements: We thank Kevin Audi, Evan DeFilippis, Felipe Gonzalez, Leah Luben, and especially Michael Walker for excellent research assistance. All errors remain our own.
* Corresponding author.

Suggested Citation: Hicks, JM, Kremer, M and Miguel, E 2014. *Estimating deworming school participation impacts and externalities in Kenya: A Comment on Aiken et al. (2014)*, Original author response to 3ie Replication Paper 3, part 1. Washington, DC: International Initiative for Impact Evaluation (3ie).

1 Executive Summary

Aiken et al. (2014) undertake a replication of Miguel and Kremer (2004), which evaluates a Kenyan project in which mass treatment with deworming drugs was randomly phased into schools, rather than to individuals, allowing estimation of overall effects even in the presence of epidemiological effects due to reduced transmission of disease. We thank Aiken et al. for undertaking this work and are pleased to be part of a continuing conversation regarding the health and development impacts of school-based deworming. We are supportive of the process of replication as a normal part of scientific research, and have been active supporters of growing efforts to promote greater transparency and reproducibility in the social sciences (Miguel et al., 2014).

This document comments on the replication analysis presented in Aiken et al. (2014). The tables in Aiken et al. (2014) confirm the main empirical findings of the Miguel and Kremer (2004) paper, namely 1) that deworming creates positive epidemiological externalities, which implies that individually randomized studies will underestimate the impact of deworming; and 2) that deworming increases school participation.

In particular, Aiken et al. (2014, Table 10) find substantial epidemiological externalities on worm infections among untreated classmates (P-value < 0.05), and externalities on worm infections among schools within 0-3 km (P-value < 0.05). With regard to school participation, Aiken et al. (2014, Table 14) find externalities on school participation among classmates (P-value < 0.01), and externalities on school participation in neighboring schools within 3 km (P-value < 0.10). Aiken et al. (2014) also find that deworming increases school participation by 5.7 percentage points in treatment schools relative to control schools (Table 18, P-value < 0.01); the comparable deworming impact on school participation in Miguel and Kremer (2004) was 5.1 percentage points (P-value < 0.01). The strong evidence of within-school externalities in Aiken et al. (2014) implies that one of the key conclusions of the Miguel and Kremer (2004) paper – that individually randomized studies of the impact of deworming will underestimate the true impact – remains valid. Aiken et al. (2014) also find that worm infections impact school participation using an instrumental variables approach (P-value < 0.05).

Aiken et al. (2014) helpfully correct a number of issues in the Miguel and Kremer (2004) paper, including: (1) a number of rounding errors in reported coefficients, some of which led to associated errors in reported P-values, and (2) some tables reporting regressions run on intermediate, rather than final, versions of data sets. These inconsistencies were introduced during the editing process when the paper was being prepared for publication, and neither of these lead to substantial changes in coefficient estimates. Aiken et al. (2014) also discuss cases of inaccurately labeled statistical significance. The effect on anemia was originally reported as significant with P-value < 0.05 but is found in re-analysis to have a P-value of 0.19. The coefficient estimate and standard error in Miguel and Kremer (2004) were reported correctly, but we believe the significance level was misreported due to a calculation of the t-statistic using rounded coefficients.

The replication also corrects an error in the original code used to estimate the externalities associated with deworming. As a result of this error, Miguel and Kremer (2004) estimate externalities for all schools 0-3 km away, and for schools 3-6 km away that are among the 12 closest schools, rather than among all schools within 3-6 km, as stated in the paper.

The externality effect on moderate-to-heavy worm infections from treated pupils attending schools 3-6 km away was statistically significant in the original Miguel and Kremer (2004) analysis, but is not significant in the updated analysis in Aiken et al. (2014). The point estimate on the 3-6 km externality term in the school participation analysis was negative but not statistically significant in the original Miguel and Kremer (2004) analysis, and remains so in the updated analysis. The fact that there are no infection externalities in the 3-6 km range (with the updated data) means there is little reason to expect school participation externalities at this distance.

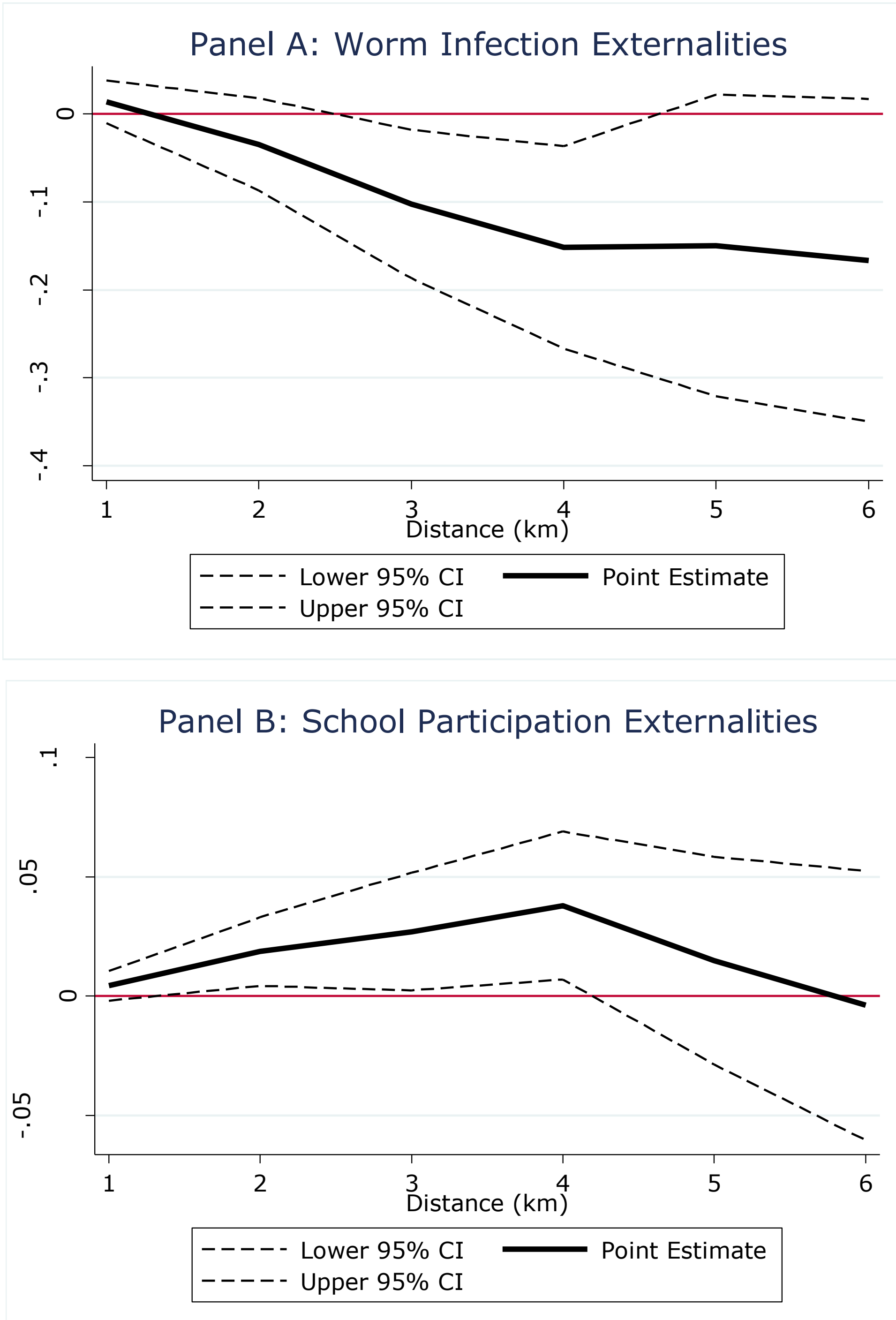
When the 3-6 km externality terms are omitted, externality effects are strong both within schools, and across schools up to 3 km away, both for worm load and for school participation. There are obviously also strong externalities when one includes the schools out to 6 km that are among the twelve closest schools, as in Miguel and Kremer (2004). Estimated overall externality effects that go out to 3 km, to 4 km, or to the 12 closest schools within 6 km are all also strong.

However, an estimator for overall externalities that goes out beyond this distance, and that puts extensive weight (due to the large numbers of schools at that distance) on the not statistically significant 3-6 km externality estimate adds large amounts of "noise" to the overall externality estimate. We demonstrate that, under reasonable assumptions, the estimator that excludes the 3-6 km externalities is preferred under the standard statistical criterion of minimizing mean squared error. We thus differ with Aiken et al. (2014) over the appropriate way to calculate overall deworming externalities on school participation and the overall impact of deworming on school participation in the updated data.

Figure 1, Panel B demonstrates how standard errors on school participation externality estimates become large when one considers schools beyond 4 km. The average cross-school externality impact on school participation is positive and statistically significant at 95% confidence at distances of 0-2, 0-3 and 0-4 km. This is evidence of deworming externalities for schools within up to 3 to 4 km of treatment schools, but not for more distant schools.

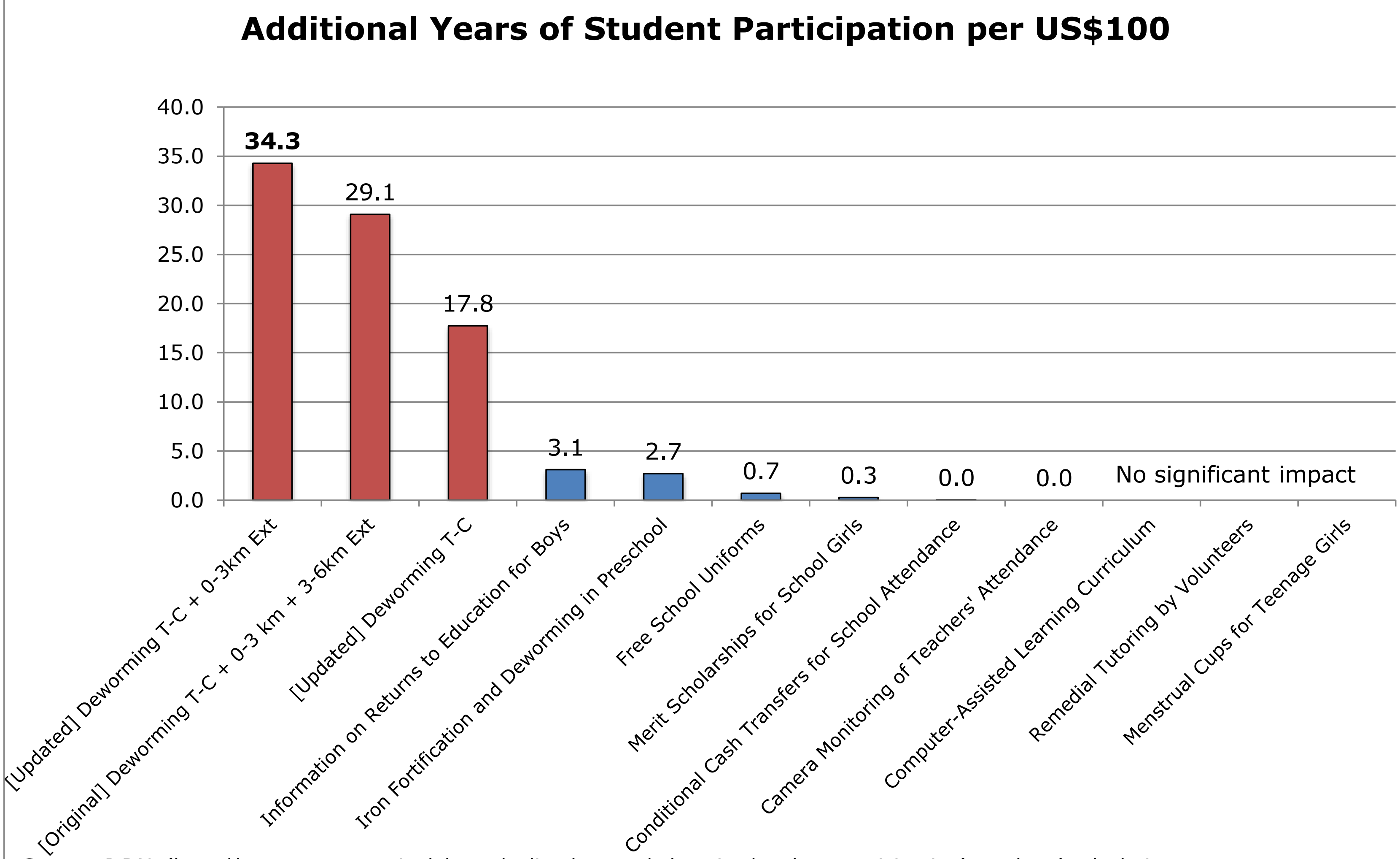
The "cost-effectiveness" of deworming in terms of boosting school participation is nearly unchanged relative to the original paper, using the updated data and considering the direct effects and the externalities up to 3 km, with 34.3 additional years of school participation per $100 of spending on deworming with the updated data (and 29.1 additional years per $100 in the original analysis). Focusing on the most conservative treatment effect estimate, the "naïve" T – C difference, also implies that deworming is a highly cost-effective approach to reducing school absenteeism in this setting, with 17.8 additional years of school participation per $100 of deworming spending, placing it among the most cost-effective interventions yet evaluated in education studies (see Figure 2).

Figure 1. Average externality impacts at various distances



Note: Panel A plots the "average externality effect" estimates presented in Table 3 (for worm infections) and Panel B plots the "average externality effect" estimates from Table 4 (for school participation). See the notes to these tables for details on the regressions.

Figure 2: Cost-effectiveness of school participation interventions



Source: J-PAL (http://www.povertyactionlab.org/policy-lessons/education/student-participation), authors' calculations.
Notes: Some values are adjusted for inflation but the deworming costs are not. The cost of deworming is US$ 0.49, as in Miguel and Kremer (2004). This is likely an overestimate, since the per-child cost in the 2009 Kenya National School-Based Deworming Program was US$0.36 (http://www.dewormtheworld.org/our-work/kenya-national-school-based-deworming-program for more details).

New evidence is rapidly accumulating on the educational and socio-economic impacts of child deworming. A key lesson of Miguel and Kremer (2004) is that traditional individual-level randomized designs will miss any spillover benefits of deworming treatment, and this could contaminate estimated treatment effects. Thus cluster randomized designs provide better evidence. Three new working papers with such cluster randomized designs estimate long-run impacts of child deworming up to 10 years after treatment; these effects on long-run life outcomes are arguably of greatest interest to public policymakers.

Croke (2014) finds positive long-run educational effects of a program that dewormed a large sample of 1 to 7 year olds in Uganda, with statistically significant average test score gains of 0.2 to 0.4 standard deviation units on literacy and numeracy 7 to 8 years later. The Ugandan program is one of the few studies to employ a cluster randomized design, and earlier evaluations of the program had found large short-run impacts on child weight (Alderman et al., 2006; Alderman, 2007). Croke (2014, p. 16) also surveys the emerging deworming literature and concludes that "*the majority of clustered trials show positive effects*".[1]

Two other new working papers explore the long-run impacts of the Kenya program we study. While the primary school children in the Miguel and Kremer (2004) sample were probably too old for deworming to have major impacts on brain development, and there was no evidence of such impacts, Ozier (2014) estimates cognitive gains 10 years later among children who were 0 to 2 years old when the deworming program was launched and who lived in the catchment area of a treatment school. These children were not directly treated themselves but could have benefited from the positive within-community externalities generated by mass school-based deworming. Ozier (2014) estimates average test score gains of 0.3 standard deviation units, which is equivalent to roughly half a year of schooling and similar to the effect magnitudes estimated by Croke (2014). This provides further strong evidence for the existence of large, positive, and statistically significant deworming externality benefits within the communities that received mass treatment.

Finally, Baird et al. (2014) followed up the Kenya deworming beneficiaries from the Miguel and Kremer (2004) study during 2007-2009 and find large improvements in their labor market outcomes. Ten years after the start of the deworming program, men who were eligible to participate as boys work 3.5 more hours each week, spend more time in entrepreneurship, are more likely to hold manufacturing jobs with higher wage earnings, and have higher living standards. Women who were eligible as girls have better educational outcomes (including higher rates of passing the primary school completion exam and enrolling in secondary school), are more likely to grow cash crops, and reallocate labor time from agriculture to entrepreneurship. The impacts of subsidies on labor hours are sufficiently large that the net present value of government revenue generated by deworming subsidies exceeds the cost of the subsidies, creating an "expenditure Laffer effect". In the preferred estimate, each additional $1 in child deworming subsidies increases the net present value of government revenue by $13.

Taken together, and building on Miguel and Kremer (2004), Alderman et al. (2006), and Alderman (2007), this new wave of studies promises to bring considerable new

[1] One exception is Awasthi *et al.* (2013), who use a clustered randomized design and find positive, but not statistically significant, effects of deworming on infant mortality and weight in a lightly infected preschool population in India. This study does not track later educational or labor market outcomes.

evidence to bear on the long-run impacts of childhood deworming on important life outcomes in areas with high worm infection rates.

We focus on the most important technical issues of Aiken et al.'s (2014) replication analysis in Section 2, and address additional points raised in their report in Section 3. In Appendix A, we present all of the tables from the original Miguel and Kremer (2004) paper, updated using the final data and correcting any coding errors discussed in Aiken et al. (2014), and in Appendix B we present our preferred final tables using the updated data. The tables we present in this note should be considered the fully "updated" version of the analysis in the 2004 paper, and these may be of interest to scholars, non-profit organizations, and policymakers. The full replication dataset, code, and documentation are available from the authors, and we welcome further analysis by other interested researchers.

2 Technical response to Aiken *et al*. (2014)

In this section, we first provide an overview of the Miguel and Kremer (2004) study, and then go on to discuss the cross-school externality findings and other issues raised in Aiken et al. (2014).

2.1 Background on Miguel and Kremer (2004)

It is useful to briefly summarize Miguel and Kremer (2004)'s approach and findings up front. The abstract to the paper summarizes its main goals, results and contributions, and we reproduce it here:

> *"Intestinal helminths—including hookworm, roundworm, whipworm, and schistosomiasis—infect more than one-quarter of the world's population. Studies in which medical treatment is randomized at the individual level potentially doubly underestimate the benefits of treatment, missing externality benefits to the comparison group from reduced disease transmission, and therefore also underestimating benefits for the treatment group. We evaluate a Kenyan project in which school-based mass treatment with deworming drugs was randomly phased into schools, rather than to individuals, allowing estimation of overall program effects. The program reduced school absenteeism in treatment schools by one-quarter, and was far cheaper than alternative ways of boosting school participation. Deworming substantially improved health and school participation among untreated children in both treatment schools and neighboring schools, and these externalities are large enough to justify fully subsidizing treatment. Yet we do not find evidence that deworming improved academic test scores."*

Miguel and Kremer (2004) evaluate a deworming program conducted by the non-governmental organization ICS in 75 Kenyan primary schools. Schools were divided into three groups of 25 schools each, and these groups were phased into deworming treatment over time, thus allowing the data to be analyzed using stepped-wedge methods. Deworming treatment began in March 1998 among the 25 Group 1 schools, and took place between March and June 1999 for both Group 1 and Group 2 schools; Group 3 schools did not receive deworming treatment in either of these two years.

It is worth reviewing the nature of disease transmission since these bear on the potential for epidemiological externalities. Geohelminths are deposited in stool, and while adults in the area typically use latrines, children are more likely to defecate in the open. This can lead to transmission of geohelminths when children defecate near their school or home. Schistosomiasis involves transmission through fresh water (via intermediate hosts) and in the study area can be transmitted when children travel to Lake Victoria to bathe or fish. It is thus likely to be transmissible over somewhat larger distances than geohelminths, particularly as part of the life cycle of the parasite occurs in snails and the snails themselves are mobile. Treatment for geohelminths was provided in all treatment schools, while treatment for schistosomiasis was only provided in those schools with sufficient prevalence of the disease, typically in schools that were located near Lake Victoria.

It was only after evidence of externalities among untreated children in the treatment schools, both in terms of worm infections and school attendance, was detected, that the decision was made to investigate the existence of externalities across neighboring schools. This analysis initially focused on the schools closest to the treatment schools. Finding evidence for positive deworming treatment effects on both worm infections and school participation at those distances, impacts were then estimated at even greater distances from each school. Externality results were presented up to 6 km away from each school, and no farther, not because there were *a priori* reasons to expect effects at 6 km *ex ante*, but rather because having found effects at 3 km – and knowing that effects could be biased downward if spillover effects were not included – we thought it worth checking for effects further out, as long as they could be estimated with sufficient precision. Note that the key test in Miguel and Kremer (2004) for the existence of externality effects lies in the statistical significance of externalities at various distances, rather than being based on a weighted sum of these externalities.

2.2 Results common to Aiken et al. (2014) and Miguel and Kremer (2004)

Miguel and Kremer (2004) conclude that deworming reduced worm infections and improved school participation in Kenyan primary schools, when deworming treatment schools are compared to control schools that did not receive deworming drugs. The paper also finds evidence of large externality (spillover) benefits in these two dimensions among untreated children (those who did not receive deworming drugs) in treatment schools. It presents evidence for large externality benefits on worm infections for those attending other schools located near treatment schools (within 0 to 3 km) and for those located farther away from treatment schools (3 to 6 km away). It presents evidence for large externality benefits on school participation within 3 km of treatment schools, but finds no statistically significant externality effect from 3-6 km away.

The Aiken et al. (2014) replication report affirms most of these findings in the Miguel and Kremer (2004) paper. Epidemiological externalities on worm infections within schools, and across schools located up to 3 km away remain strong. Direct effects of deworming on school participation and externality effects within schools remain strong. As in Miguel and Kremer (2004), there are no statistically significant externality effects on school participation beyond 3 km. As in Miguel and Kremer (2004), there is no statistically significant effect on test scores within the time period examined.

However, the replication was useful in highlighting some discrepancies, and we thank the replication team for enabling us to jointly update the scientific record. A key difference is one of interpretation of the cross-school externalities on school participation. We interpret the results as indicating statistically significant externalities at 0-3 km and no statistically significant effects at 3-6 km. Aiken et al. (2014) note that the confidence interval on a weighted sum of the two coefficients (with weights given by the average number of schoolchildren at each distance) includes zero, and therefore conclude that there are no cross-school externalities on school participation.

2.3 Errors and discrepancies addressed in Aiken et al. (2014)

Aiken et al. (2014) helpfully re-analyze the data in Miguel and Kremer (2004), and discuss a number of errors. We review these below, starting with rounding errors and minor changes to the data set (which accounted for the majority of the discrepancies), and then considering the coding error that led to measurement of externalities in the 3-6 km range only among schools that were among the 12 closest to the reference school.

2.3.1 Rounding errors and data updates

A leading reason for these errors had to do with the rounding of some figures after reducing the number of significant figures from three to two (for aesthetic reasons) during the journal revision process. For instance, a figure of 0.7745 was initially presented as 0.775 in tables, but then incorrectly rounded up to 0.78 (rather than down to 0.77) when we moved to presenting only two digits in the published version of the paper. By definition, rounding errors are small in magnitude, and they lead to only small changes in the results.

Aiken et al. (2014) additionally discuss several cases of inaccurately labeled statistical significance. We believe that some of these were also the result of rounding in coefficient estimates and standard errors, which led to inaccuracies in t-statistics. Some of these led to results reported as at traditional levels of confidence becoming insignificant. The most important among these is that presented in Miguel and Kremer (2004), Table V – "Proportion anemic", which was originally reported as statistically significant with 95% confidence, but is found in reanalysis to have a p-value of 0.19. Note that the coefficient estimate and standard error in the original Miguel and Kremer (2004) paper were reported correctly, so the magnitude of the effect is unchanged at -2 percentage points, but the statistical significance level was misreported. While it was important to include an examination of anemia from a medical perspective, Miguel and Kremer (2004) note that anemia is not likely to be a main channel of impact in the setting examined because only 4% of the population was anemic. Correspondingly, this is not one of the major findings of the original paper.

A second reason for these errors is that intermediate versions of several datasets were used in production of the paper, and not all of the tables were fully updated with final versions of the data during the journal revision process. This accounts for the largest number of discrepancies with the original paper. However, the extent of final data cleaning was only moderate over that time, so that using different versions of the data leads to very similar results.[2] We support the growing trend among journals to require authors to prepare

[2] Data cleaning, in both Kenya and the United States, was an ongoing process on these large, original data sets during 1998-2002, and this led to the existence of various "intermediate" versions of data, versions that were progressively cleaner over time. Cleaning typically took the form of eliminating duplicate observations, correcting data entry errors through hard copy checks, and better matching

online replication data materials prior to publication, since we believe that this will make it less likely that these sorts of errors will happen going forward.

Aiken et al. (2014) note at several points that the changes in results due to these rounding errors and data updates are generally small (in the range of 0.01 for many estimates), and do not substantively change the results in Miguel and Kremer (2004).

2.3.2 Externality effects 3 to 6 km away from treatment schools

The biggest issue is an error in the construction of the local population density terms at a radius of 3-6 km from each school. This error meant that whereas Miguel and Kremer (2004) reported externalities between 3-6 km away from a school, it actually measured externalities only for those schools within 3–6 km that were among the 12 closest schools.[3] Some deworming treatment effects are marginally larger in magnitude and somewhat more precisely estimated when all schools within 3-6 km are included, and some are smaller or less precisely estimated. There are a few noteworthy changes and we focus on those here.

This issue did not affect the construction of the 0-3 km externality terms, but in a number of cases it did affect the construction of the 3-6 km externality terms. In no case did a school have more than 12 schools within a 4 km radius, so externality terms up to that radius were correct. Three quarters of schools had twelve or fewer schools within 5 km. However, at distances greater than 5 km many schools are affected. Arguably, one should not expect to find substantial spillover effects from schools that were not among the 12 closest neighboring schools, and including these more distant schools in the measured externalities naturally drives the average externality effect towards zero.

It is worth noting that the approach in the original Miguel and Kremer (2004) paper still produces a well-defined statistic, i.e., an externality measure that focuses on up to the 12 closest schools. In fact, many influential recent empirical explorations of social effects employ related measures, for instance, measures of social networks that restrict attention to an individual's 10 or 15 "closest" acquaintances (see for instance, Conley and Udry, 2010). Hence the use of this statistic is still meaningful in assessing the presence of externalities, but it does of course have a different interpretation than the one provided in the original paper.

Once all schools within a 3-6 km radius are included, rather than just the 12 closest schools, Aiken *et al.* (2014) find direct effects (namely, the Treatment vs. Control difference) and within-school externality impacts for worm infections that are marginally larger in magnitude than the original study. Furthermore, the replication confirms Miguel and Kremer (2004)'s findings of cross-school epidemiological externality impacts within 3 km, as well as the direct effects and within-school externality impacts for school attendance. It is mainly the cross-school externality estimates beyond 3 km that are affected. With regard to worm infections, Miguel and Kremer (2004) find reductions within 3-6 km, but this

across files. Economics journals ask authors for specific revisions, and in revising the paper we also discovered and corrected minor errors in our dataset. However, we did not systematically update all of the other tables, so different tables in Miguel and Kremer (2004) were based on slightly different versions of the dataset.

[3] There was a second, and much more minor, error in the construction of the externality measures, which affected only two schools. We explain this error in detail in Section 3.

finding is not statistically significant upon re-analysis.[4] There is no evidence of externality effects on school participation among the full set of schools within 3–6 km.

The standard errors on the "overall" 3-6 km externality effect become much larger, nearly doubling in the worm infection case and more than doubling in the estimation of the average effect of school attendance externalities. Including all schools, instead of only the nearest twelve, is what adds "noise" to the estimated overall 3-6 km externality effects. With such large standard errors, the degree of noise in the estimates of overall externalities becomes very large, and the estimates are relatively uninformative about the underlying signal in the data.

Note that the 3-6 km externality effect for school participation was not statistically significant in the original Miguel and Kremer (2004) paper. At a distance over which overall externalities can be precisely estimated (up to 3 km), the main finding remains that there are large and highly significant cross-school externalities for both worm infections and school attendance. Using the updated data, the estimated average cross-school externality effect of deworming on worm infections is a reduction of 10.2 percentage points (s.e. 4.3, P-value < 0.05), shown in column 2 of Table 1. The estimated average cross-school externality effect of deworming on school participation is a gain of 2.7 percentage points (s.e. 1.3, P-value < 0.05), shown in column 2 of Table 2.

Aiken et al. (2014) follow the original paper in focusing on externalities out to 6 km, and calculate the "overall effect" of deworming on school attendance by taking the weighted sum of the two coefficients (on 0-3 km and 3-6 km, with weights given by the average number of schoolchildren at each distance). The weight given to the 3-6 km externality term increases substantially once all schools in the 3-6 km range are included, rather than just those among the closest 12. The authors go on to conclude: *"there was no evidence of a between-school indirect effect or an overall effect from the intervention on school attendance"* (p. iii).

We disagree with this claim, and believe it is a misinterpretation of the statistical evidence presented in their tables. Given the updated data, a regression specification different from that in the original paper is necessary to precisely estimate the overall externality effect of deworming. While it is natural to first replicate the exact specification used in the original paper, the changes to the data mean that this estimator is no longer appropriate. More reliable conclusions can be reached by excluding the 3-6 km externality effect from the calculation of overall effects, since it is adding a tremendous amount of "noise" to the estimate.

Miguel and Kremer (2004) demonstrate that the "naïve" mean difference between Treatment and Control units, what we call the T-C difference, underestimates the total impact of treatment in the presence of epidemiological externalities and propose a simple and tractable methodology for estimating cross-unit externalities. The idea behind the estimation strategy in Miguel and Kremer (2004) is that the "naïve" T-C difference – and in fact any estimator that only considers externalities up to a certain distance away from each school – would serve as a lower bound on the true overall impact of deworming due to the presence of positive spillovers.

[4] However, there is evidence that these longer-range 3-6 km externalities exist for schistosomiasis infection, as shown in Table of Aiken et al. (2014) and Table VII of Appendix A below, but schistosomiasis drugs were given in only a minority of schools (where the disease was common).

The original paper presented externality results up to 6 km away from each school, and no farther, not because we had conceived of this exact test *ex ante*, but because we could not precisely estimate overall externality effects at greater distances. Page 186 of the original paper explains why we chose to focus on externality impacts out to 6 km from each school – but not beyond – at that time:

> *"Due to the relatively small size of the study area, we are unable to precisely estimate the impact of additional treatment school pupils farther than six kilometers away from a school, and thus cannot rule out the possibility that there were externalities at distances beyond six kilometers and possibly for the study area as a whole, in which case the estimates presented in Table VII (and discussed below) would be lower bounds on actual externality benefits."*

However, the effect of the variable construction issue was that instead of measuring externalities between 3-6 km, we were in fact measuring externalities over a narrower range (typically a subset of schools within the 3-6 km range). The key issue that arises when we expand the measure of externalities to all schools within 3-6 km is that the precision of the overall externality estimate goes down dramatically. There is a natural statistical interpretation for this reduction in precision using the updated data. The externality coefficient estimates are multiplied by the average number of treatment pupils in the appropriate range (either 0-3 km or 3-6 km), and this number increases dramatically in the updated data that includes all schools in the 3-6 km range. Since the updated 3-6 km externality terms are not statistically significant for worm infections (Table 1, column 3) or school participation (Table 2, column 3), this means that a lot of "weight" in the calculation of the overall externality effect is placed on distant schools with an imprecisely estimated "zero" externality effect.[5]

As shown in Table 1, the standard error on the average overall 3-6 km externality effect nearly doubles in the estimation of infection externalities; you can see this in Table 1 by comparing the standard error of 0.042 in column 6 (results from the original paper, with the coding error) to the standard error of 0.079 in column 3 (results using the updated and corrected data). Similarly, it more than doubles in the estimation of school participation effects (comparing the standard error of 0.011 in column 6 to the standard error of 0.024 in column 3 of Table 2). This marked reduction in statistical precision is also clear visually in Figure 3, where the 95% confidence intervals increase substantially once the updated 3-6 km externality effects are included, for both infection outcomes and school participation outcomes. These large confidence intervals are relatively uninformative, and also lead the estimate of total deworming impacts to be much less precisely estimated.

If we impose a sensible decision rule and exclude externality estimates that are simply too imprecisely estimated to be informative (as we did in the original analysis), then including the 3-6 km effect is inappropriate with the updated data. The best way to think about it is that including these 3-6 km externalities is like adding a very "noisy zero" estimate to what is otherwise quite a precise estimate. It is appropriate to focus on the estimator that includes the "naïve" treatment minus control difference plus the 0-3 km

---

[5] A second issue is that, while more data is utilized by bringing in all schools between 3-6 km, precision falls because there is relatively less idiosyncratic variation in the number of treatment school pupils (relative to total pupils) in larger geographic areas.

externalities, since these are both precisely estimated, and these together constitute a lower bound on the overall effect of deworming under the reasonable assumption that deworming externality effects are non-negative. Even focusing on the precisely estimated "naïve" estimator – the simple T minus C difference – which is downward biased since it excludes all cross-school externality effects, would be preferable to employing the estimator that incorporates externalities from 3-6 km, since the naïve estimator is precisely estimated and provides a lower bound on the magnitude of the true effect.

It is useful to think about including additional externality estimates in terms of the usual goal of choosing an estimator that minimizes "mean squared error". Recall that mean squared error is the sum of the variance of an estimator plus the square of its bias. Including further externality terms in the analysis helps reduce bias in the estimation of the overall effect (by capturing more of the externalities) but the analyst faces a trade-off if their inclusion increases the variance of the resulting estimator. In cases where standard errors increase dramatically with the inclusion of additional terms, means squared error is reduced by focusing on precisely estimated effects that constitute a lower bound on the true overall effect. Aiken et al.'s (2014) conclusion that there is no significant evidence of a deworming effect on school participation is driven by their decision to take a precisely estimated effect that is a lower bound on the true impact – and indicates large school participation gains – and add lots of "noise" to it, by including the 3-6 km externality effects. In our view, this is a statistically inappropriate approach given the updated data.

The patterns in the tables illustrate this point. Using the original data, including the 3-6 km externality effect in the overall deworming effect does not appreciably increase the standard error on the overall effect: in Table 1, the standard error remains unchanged at 0.055 when the 3-6 km term is included in the worm infection analysis (as shown in the bottom row of columns 5 and 6), and similarly the standard error on the overall effect remains nearly unchanged in the school participation analysis (comparing columns 5 and 6 of Table 2). With the original data, there does not appear to be much of a trade-off between bias and statistical precision at all. Moreover, with the original data the 3-6 km externality effect is statistically significant on its own (Table 1, column 6), so it is natural to include it in the calculation of overall effects. While the 3-6 km externality effect is not significant for school participation using the original data (Table 2, column 6), it is reasonable to consider the possibility that there might be schooling externalities at that distance, given the worm infection externality gains at 3-6 km.

(Note that in the original working paper version of the paper (Miguel and Kremer, 2001), we did not consider the 3-6 km externality effects in our calculation of overall deworming impacts on school participation since they were not statistically significant, and in fact we did not even present them in the analysis (in Table 11). During the paper revision process at the journal *Econometrica*, we later incorporated the 3-6 km externality effects into the school participation regressions to maintain analytical consistency with the infection externality regressions, and given the existence of statistically significant 3-6 km worm infection effects using that data.)

Table 1: Worm infection results from Miguel and Kremer (2004), updated and original

	UPDATED			ORIGINAL		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Treatment Indicator	-0.347***	-0.333***	-0.313***	-0.347***	-0.311***	-0.247***
	(0.052)	(0.052)	(0.057)	(0.052)	(0.052)	(0.053)
Treatment pupils w/in 3 km		-0.234**	-0.212**		-0.249***	-0.256***
(per 1000 pupils)		(0.097)	(0.104)		(0.085)	(0.087)
Treatment pupils w/in 3 - 6 km			-0.050			-0.140**
(per 1000 pupils)			(0.077)			(0.060)
Total PSDP 'eligible' students w/in 3 km		0.069*	0.046		0.074**	0.109***
(per 1000 pupils)		(0.037)	(0.036)		(0.033)	(0.040)
Total PSDP 'eligible' students w/in 3-6 km			-0.022			0.133**
(per 1000 pupils)			(0.039)			(0.056)
School average of mock score, 1996	-0.208***	-0.216***	-0.188***	-0.208***	-0.220***	-0.093
	(0.055)	(0.052)	(0.073)	(0.055)	(0.048)	(0.068)
Calculated Effects						
Average 0-3 km externality effect		-0.102**	-0.090**		-0.111***	-0.106***
		(0.043)	(0.044)		(0.038)	(0.037)
Average 3-6 km externality effect			-0.052			-0.096**
			(0.079)			(0.042)
Average overall cross-school externality effect		-0.102**	-0.146		-0.111***	-0.212***
		(0.043)	(0.110)		(0.038)	(0.065)
Overall deworming effect	-0.347***	-0.435***	-0.459***	-0.347***	-0.421***	-0.460***
	(0.057)	(0.061)	(0.091)	(0.057)	(0.055)	(0.055)

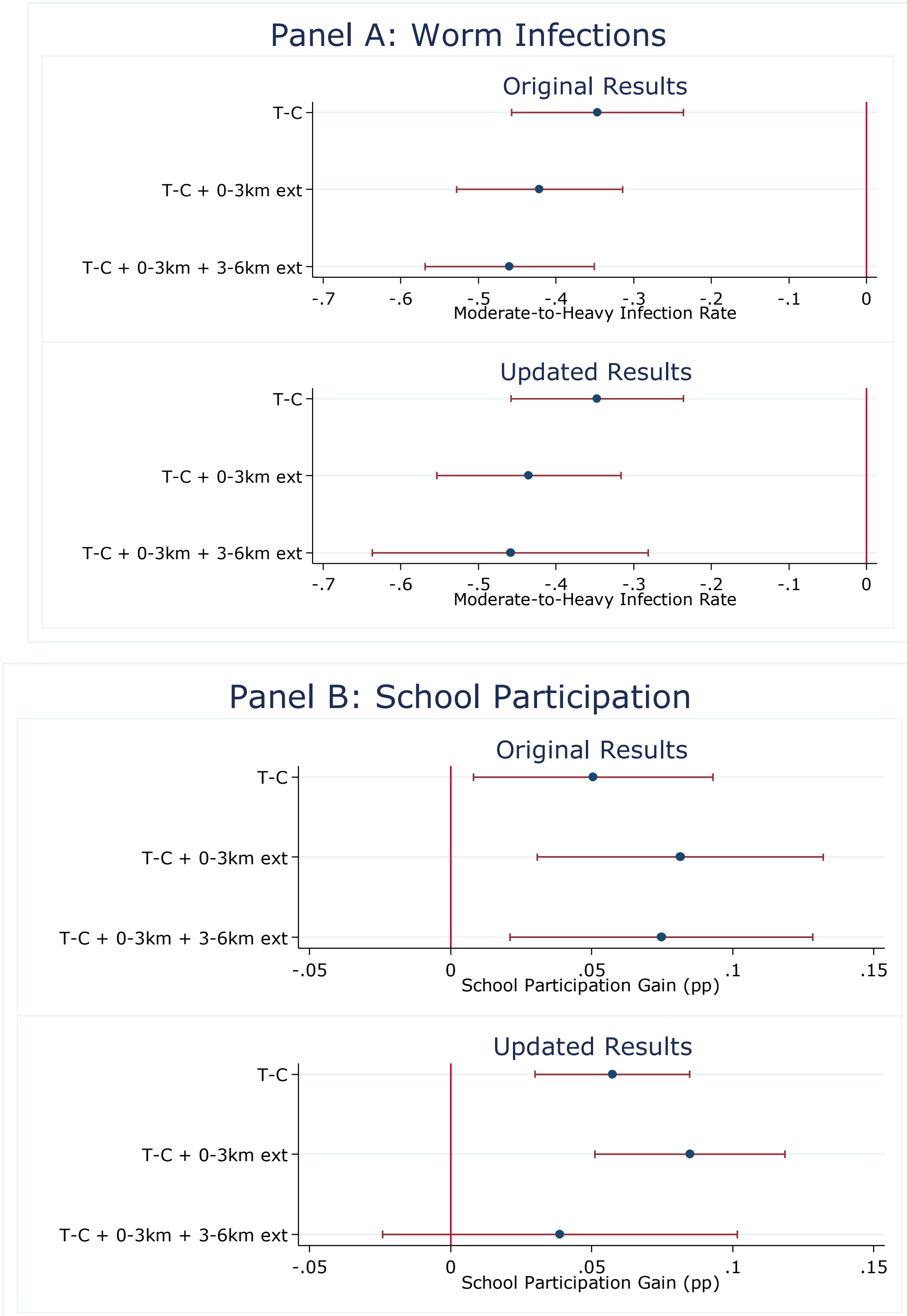
Note: The sample size in columns (1)-(3) is 2,330, and in (4)-(6) is 2,328. The sample includes pupils in grades 3–8, in 1999 Group 1 and Group 2 schools. Results are from probit estimation, where observations are weighted by total school population. The dependent variable is an indicator for moderate-to-heavy infection. Eligible pupils include girls less than 13 years old and all boys. Additional explanatory variables include indicators for 1998 grade and school SAP participation. Robust standard errors are in parentheses, and disturbance terms are clustered within schools. Stars denote statistical significance at 99 (***), 95 (**), and 90 (*) percent confidence.

Table 2: School participation results from Miguel and Kremer (2004), updated and original

	UPDATED			ORIGINAL		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Treatment Indicator	0.057***	0.058***	0.055***	0.051**	0.054**	0.055**
	(0.014)	(0.014)	(0.014)	(0.022)	(0.023)	(0.023)
Treatment pupils w/in 3 km		0.045**	0.038*		0.046**	0.048**
(per 1000 pupils)		(0.021)	(0.021)		(0.018)	(0.019)
Treatment pupils w/in 3 - 6 km			-0.024			-0.013
(per 1000 pupils)			(0.015)			(0.015)
Total PSDP 'eligible' students w/in 3 km		-0.030**	-0.030**		-0.031***	-0.037***
(per 1000 pupils)		(0.013)	(0.012)		(0.012)	(0.012)
Total PSDP 'eligible' students w/in 3-6 km			0.012			-0.014
(per 1000 pupils)			(0.009)			(0.012)
School average of mock score, 1996	0.071***	0.071***	0.078***	0.063***	0.064***	0.055***
	(0.021)	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)
Calculated Effects						
Average 0-3 km externality effect		0.027**	0.023*		0.028**	0.029**
		(0.013)	(0.013)		(0.011)	(0.012)
Average 3-6 km externality effect			-0.040			-0.009
			(0.024)			(0.011)
Average overall cross-school externality effect		0.027**	-0.017		0.028**	0.020
		(0.013)	(0.030)		(0.011)	(0.013)
Overall deworming effect	0.057***	0.085***	0.039	0.051**	0.081***	0.075***
	(0.014)	(0.017)	(0.032)	(0.022)	(0.026)	(0.027)

Note: The sample size in columns (1)-(3) is 56,496, and in (4)-(6) is 56,487. The dependent variable is average school participation in each year (Year 1: May 1998 - March 1999; Year 2: May 1999 - November 1999). Participation is computed among all pupils enrolled at the start of the 1998 school year; pupils present during an unannounced NGO school visit are considered participants. Additional controls include an indicator for girls < 13 years and all boys; the rate of moderate-heavy infections in geographic zone, by grade (zonal infection rates among grade 3 and 4 pupils are used for pupils initially recorded as drop-outs; rates among grade 5 and 6 pupils are used for grades 5 and 6, and similarly for grades 7 and 8); 1996 school average test score; indicators for participation in the SAP, alone and interacted with an indicator for 1998; indicators for 1998 grade of pupil; and indicators for semester of observation. Robust standard errors are in parentheses, and disturbances are clustered within schools. Stars denote statistical significance at 99 (***), 95 (**), and 90 (*) percent confidence.

Figure 3: Original vs. updated "overall effect", with 95% confidence intervals



Note: Panel A displays the "overall effect" of deworming, as calculated in the bottom panel of Table 1 (for worm infections) and Panel B displays the "overall effect" of deworming from Table 2 (for school participation). See the notes under these tables for details on the regressions.

In contrast, the pattern of results using the updated data indicates that it is not appropriate to include the 3-6 km externality effects in the calculation of overall deworming impacts. First, the 3-6 km externality effect is not statistically significant for either worm infections (with a coefficient estimate of -0.050 and standard error of 0.077 implying a P-value of 0.52, in Table 1 column 3), or for school participation (Table 2, column 3). Second, there is a tremendous loss of statistical precision in the overall effect estimate when 3-6 km externality effects are included in the calculation. For worm infections, the standard error on the overall effect estimate increases by 50% (from 0.061 to 0.091, Table 1 columns 2 and 3) when the 3-6 km externality effect is included. For school participation, the standard error on the overall effect estimate nearly doubles, from 0.017 to 0.032 (Table 2, columns 2 and 3). This doubling of the standard error in the school participation analysis is equivalent to increasing the variance of the estimator roughly four-fold, so the reduction in bias from including the 3-6 km externality effect would have to be very large to justify its inclusion under the criterion of minimizing the mean squared error (MSE). Yet it is unlikely that the 3-6 km externality effect on school participation is substantial given the lack of worm infection externality impacts at 3-6 km.

Some straightforward calculations suggest that the estimator that excludes the 3-6 km externality terms from the calculation of overall deworming impacts on school participation is preferable under the criterion of minimizing MSE. In particular, we show that the increase in MSE due to additional noise from including the 3-6 km term is likely to be more than six times greater than any decrease in the MSE due to reducing bias.

To see this, define the estimator that includes the Treatment minus Control effect plus the 0-3 km externality effect as β_1 (this is the estimate presented in the bottom row of Table 2, column 2), and the estimator that also includes the 3-6 km externality effect as β_2 (column 3). An estimate of the variance of β_1 is the square of its standard error, or 0.017^2, and similarly for the variance of β_2 (0.032^2). For simplicity, we conservatively assume that Bias(β_2) = 0, in other words, all deworming externality effects are captured within 6 km. The estimator that excludes the 3-6 km externality terms is preferred under the mean squared criterion – in other words, MSE(β_1) < MSE(β_2) – as long as Bias$(\beta_1)^2 < (0.032^2 - 0.017^2) = 0.000735$, or equivalently, if Bias(β_1) – Bias(β_2) < $(0.000735)^{1/2} = 0.027$.

Recall that the direct effect of being in a treatment school is to reduce moderate-heavy worm infections by -0.31 (Aiken et al.'s Table 10). Note also that they estimate that the direct effect of being in a treatment school is to increase school participation by 0.057 (Table 14). This suggests that every percentage point reduction in moderate-heavy infection increases school participation by roughly (0.057)/(0.31) = 0.184 percentage points. Aiken et al.'s Table 10 also shows that the point estimate of the reduction in worm infections within 3-6 km is only 0.050, which is not statistically significant. This implies that the predicted gain in school participation per 1000 treated pupils within 3-6 km is approximately 0.05 * 0.184 = 0.0092, and the average externality effect (given the average number of treated pupils within 3-6 km) is 0.011. This is the predicted change in bias from including the 3-6 km externalities in the estimation of overall deworming impacts, Bias(β_1) – Bias(β_2). It is immediate that Bias(β_1) – Bias(β_2) = 0.011 < 0.027, and thus that the estimator excluding the 3-6 km externality term is preferred under the MSE criterion.

This means that the predicted decrease in MSE due to reduced bias is 0.011^2 = 0.000115. Recall from above that the increase in MSE from the additional "noise" contributed by including the 3-6 km externality effect is $(0.027)^2$ or 0.000735. Hence the increase in MSE due to the extra "noise" from including the 3-6 km externality term (0.000735) is likely to be approximately 6.4 times larger than the predicted decrease in MSE due to reduced bias (0.000115).

Even if one makes the far weaker assumption that the overall externality effect on school participation at 3-6 km is simply equal to or smaller than that from 0-3 km, one reaches the same conclusion that MSE decreases when the 3-6 km externality term is excluded. Recall from Table 2, column 2 that the overall 0-3 km externality effect on school participation is also (coincidentally) 0.027. Thus the estimator that excludes the 3-6 km externality effects (β_1) has a smaller mean squared error if the overall externality effect at 3-6 km is smaller than the 0-3 km effect. This is a very natural "monotonicity" assumption given the nature of worm transmission and reinfection, which tend to be locally concentrated and should fall at greater distances from a treatment school.

The comparison of columns 2 and 3 in Table 2 further illustrates this point. The total estimated effect incorporating the "naïve" treatment minus control difference plus the 0-3 km effect is 0.085 (s.e. 0.017), significant at 99% confidence. The total estimated effect incorporating externalities out to 6 km has a standard error of 0.032, nearly twice as large as the standard error only considering externalities out to 3 km. Regarding the negative 3-6 km point estimates, there is no obvious epidemiological reason to our knowledge why the 3-6 km effects on school participation would be negative, especially given the large, positive and significant externality effects we estimate both within-schools and within 3 km of treatment schools. We instead believe the negative and very far from statistically significant point estimates on the 3-6 km school density are most likely to be "noisy zeros", as mentioned above. It is worth mentioning again that even in the original Miguel and Kremer (2004) paper the 3-6 km externality effect on school participation was not statistically significant, but this "zero" effect becomes considerably noisier with the updated data.

In fact, once the 3-6 km variable construction is corrected, the "naïve plus 0-3 km" effect is nearly unchanged for worm infections (comparing the column 2 and column 5 results at the bottom of Table 2), and the school participation effect is slightly larger in the updated case with a somewhat smaller standard error than in the original estimation. Both the infection and school participation effects are large in magnitude and statistically significant at over 99% confidence considering externalities out to 3 km (see column 2 of Table 1 and column 2 of Table 2). Thus there remains considerable evidence that deworming led to reductions in worm infections and large improvements in school participation. But the effects beyond 3 km are simply too imprecisely estimated to be usefully employed in the analysis.

As noted above, the externality analysis was not pre-specified in advance of analyzing the data. Readers might be concerned about the possibility of data mining and selective presentation of analytical results, and wonder just how robust the externality results truly are. It is straightforward to show that the positive deworming externality results across nearby schools are robust to using different distances and specifications; it is not the case that the 3 km distance was "cherry-picked" from among the set of possible distances over which to estimate externality effects. For worm infections, the externality effects are statistically significant at 95% confidence at distances of both 0-3 and 0-4 km (Table 3, columns 3 and 4) and significant at 90% confidence at distances of 0-5 and 0-6 km (columns 5 and 6). Note that as one gets further away, one would expect the spillovers from any given school to be smaller, but the "overall" effect from multiplying the average spillover times the number of schools to stay constant or grow. The magnitude of the "overall" cross-school externality benefits become larger at increasing distances, although they are estimated with considerably less precision, especially beyond 4 km (Figure 1, Panel A). (Externality estimates are also imprecisely estimated for schools within 1 km from the reference school, since very few schools are located this close together.)

The same pattern is evident for school participation externalities. The impact of cross-school externalities is positive and statistically significant at 95% confidence at distances of 0-2, 0-3 and 0-4 km (Table 4, columns 2-4), and the magnitude is largest for the 4 km radius. Once again externality effects increase at larger distances, in this case up to 4 km, after which confidence intervals become considerably wider. In all of these regression specifications, the naïve effect on treatment schools is nearly unchanged, ranging between gains of 0.057 and 0.063 and is significant at over 99% confidence.

There is a simple bottom line on deworming externalities. As Aiken et al. (2014) show in their tables, there are large, positive and significant deworming externalities for worm infections and school participation within schools (i.e., for untreated pupils in the treatment schools). Externalities are large, positive and significant for both worm infections and school participation for schools up to 3 km and 4 km of treatment schools. This is all that is needed to show that cross-school externalities "exist": they need not hold at *all* distances in order to exist, they simply need to hold at *some* distances. Given the epidemiology of worm infections, it is reasonable that they would be more pronounced closer to treatment schools, as comes through in the updated data. In fact, the 3-6 km worm infection externalities which were statistically significant in the original paper when analysis was restricted to the 12 closest schools no longer come through once the variable construction error is corrected. The externality impacts on school participation at a distance of 3-6 km from treatment schools were not statistically significant in the original Miguel and Kremer (2004) analysis, and they remain not significant in the updated analysis.

There is an important alternative approach to estimating the impact of worms on school participation presented in Miguel and Kremer (2004), namely an instrumental variables (IV) approach. The IV method is attractive because it simultaneously exploits multiple sources of experimental variation, including both school treatment status and proximity to treatment schools (in both the 0-3 km and 3-6 km ranges), to identify a single impact of worm infections on school participation. This approach is not discussed in the Aiken et al. (2014) report, although the updated results are presented in their Table 14 (column 7), where they show that the estimated impact of a moderate-heavy worm infection on school participation is large and statistically significant (coefficient estimate -0.195, s.e. 0.095, P-value < 0.05). (Note that the sign here is negative since it is parameterized in terms of the effect of a moderate-heavy worm infection on school participation.) This result is robust to other specifications: we show in Appendix B Table IX (column 7) that the effect is even larger in magnitude at -0.282 (s.e. 0.111, P-value < 0.05), when only the school treatment status and externality effects within 0-3 km are used as instrumental variables. Note that these effects are statistically significant despite the use of a much smaller sample size, since this analysis relies on the year 1 individual worm infection data. This result provides a further piece of evidence that worm infections have a large and statistically significant impact on school participation in our setting.

Given these findings, we disagree with the claim in Aiken et al. (2014) that there is no evidence for cross-school externality impacts on school participation, and no overall effect of deworming on school participation.

For those interested in policy implications, the estimated overall average effect of deworming on worm infections using the finalized data is a reduction of 43.5 percentage points (s.e. 6.1, P-value < 0.01), shown in column 2 of Table 1. The estimated overall average effect of deworming on school participation is a gain of 8.5 percentage points (s.e. 1.7, P-value < 0.01), shown in column 2 of Table 2. We show in Figure 2 that the "cost-effectiveness" of deworming in terms of boosting school participation is nearly unchanged, relative to the original paper, using the updated data and considering the direct effects and the externalities up to 3 km, with 34.3 additional years of school participation per $100 of spending on deworming with the updated data (versus 29.1 additional years per $100 in the original analysis). Focusing on the most conservative treatment effect estimate, the "naïve" T – C difference, also implies that deworming is a highly cost-effective approach to reducing school absenteeism in this setting, with 17.8 additional years of school participation per $100 of deworming spending, placing it among the most cost-effective interventions yet evaluated in education studies, as shown in the figure.

Table 3. Worm infection regressions, with externalities at various radii

	(1) w/in 1 km	(2) w/in 2 km	(3) w/in 3 km	(4) w/in 4 km	(5) w/in 5 km	(6) w/in 6 km
Treatment indicator	-0.325***	-0.354***	-0.333***	-0.296***	-0.283***	-0.306***
	(0.047)	(0.051)	(0.052)	(0.057)	(0.064)	(0.056)
Treatment pupils within XX km	0.581	-0.236	-0.234**	-0.201***	-0.124*	-0.112*
(per 1000 pupils)	(0.535)	(0.180)	(0.097)	(0.077)	(0.072)	(0.063)
Total pupils within XX km	-0.248	0.110	0.069*	0.044	-0.011	-0.001
(per 1000 pupils)	(0.357)	(0.085)	(0.037)	(0.036)	(0.030)	(0.032)
Calculated Effects						
Average XX km externality effect	0.013	-0.035	-0.102**	-0.152***	-0.150*	-0.166*
	(0.012)	(0.027)	(0.043)	(0.059)	(0.087)	(0.094)
Overall effect	-0.311***	-0.389***	-0.435***	-0.448***	-0.432***	-0.472***
	(0.052)	(0.062)	(0.061)	(0.062)	(0.068)	(0.085)

Note: This table uses the fully corrected, updated data from Miguel and Kremer (2004). Regressions are as specified in Table 1, with the exception that we allow the radius at which externalities are considered to vary across the columns as indicated.

Table 4. School participation regressions, with externalities at various radii

	(1)	(2)	(3)	(4)	(5)	(6)
	w/in 1 km	**w/in 2 km**	**w/in 3 km**	**w/in 4 km**	**w/in 5 km**	**w/in 6 km**
Treatment indicator	0.061***	0.063***	0.058***	0.059***	0.058***	0.057***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Treatment pupils within XX km	0.179	0.093**	0.045**	0.034**	0.009	-0.002
(per 1000 pupils)	(0.131)	(0.037)	(0.021)	(0.014)	(0.013)	(0.013)
Total pupils within XX km	-0.117	-0.064***	-0.030**	-0.022**	-0.009	-0.002
(per 1000 pupils)	(0.109)	(0.025)	(0.013)	(0.009)	(0.009)	(0.008)
Calculated Effects						
Average XX km externality effect	0.004	0.019**	0.027**	0.038**	0.015	-0.004
	(0.003)	(0.007)	(0.013)	(0.016)	(0.022)	(0.029)
Overall effect	0.065***	0.081***	0.085***	0.097***	0.073***	0.053
	(0.014)	(0.015)	(0.017)	(0.020)	(0.024)	(0.033)

Note: This table uses the fully corrected, updated data from Miguel and Kremer (2004). Regressions are as specified in Table 2, with the exception that we allow the radius at which externalities are considered to vary across the columns as indicated.

2.4 Presentation of Non-Worm Infection Health Results

The key health outcome measure emphasized in Miguel and Kremer (2004) is the helminth infection rate. This is the most natural health outcome to focus on given the intervention. As shown above in Table 1, Miguel and Kremer (2004) find large and highly statistically significant decreases in worm infections due to deworming, and this result is unchanged upon re-analysis with the finalized data. As shown in Table V in Appendix A (which reproduces a table from Miguel and Kremer (2004) using the updated data), there are substantial decreases in worm infection rates for "any moderate-heavy infection", for hookworm infection, and for roundworm infection after one year of treatment (in fact, no concerns were noted in reporting of these results). Although point estimates suggest a substantial decline in schistosomiasis infection, the treatment effect is no longer statistically significant at traditional confidence levels; recall that treatment for schistosomiasis was only provided in the subset of schools with sufficient prevalence of the disease, typically in schools that were close to Lake Victoria, and thus overall infection levels and treatment effects are likely to be much more evident in those schools. Table 1 and Table 3 above also present evidence that there are large, positive epidemiological externalities associated with deworming across schools.

Beyond worm infection, Miguel and Kremer (2004) present six other health outcomes, including self-reported health in the past week, self-reports of being "sick often", height- and weight-for-age Z-scores, hemoglobin concentration, and proportion anemic. Aiken *et al.* (2014) focus their discussion on these latter four measures.

The height result was reported as a modest improvement in the original paper, and this result is entirely unchanged in re-analysis; Aiken et al. (2014) describe this as "weak evidence of a small benefit" (p. iii) but the results in the original paper seem subject to more positive interpretations. The weight-for-age Z-score and hemoglobin concentration outcomes were not found to be statistically significant either in the original study or in re-analysis.

Aiken *et al.* (2014) report an error in the reporting of the "proportion anemic" outcome, which is not statistically significant in the updated analysis. We thank the authors for updating the scientific record on this point. Note that the coefficient estimate on anemia in the original Miguel and Kremer (2004) paper was reported correctly (a reduction of 2 percentage points in anemia), so the magnitude of the effect and the standard error are unchanged, but the statistical significance level was misreported. We believe that this was due to a calculation of the t-statistic using the rounded coefficients. While anemia is interesting to study from a medical perspective, Miguel and Kremer (2004) paper note that anemia is not likely to be a main channel of impact in the setting examined because only 4% of the population was anemic. Correspondingly, this is not one of the major findings of the original paper. As we write on p. 174:

> *"Severe anemia is relatively rare in Busia: fewer than 4 percent of pupils in Group 2 schools (comparison schools in 1998) fell below the Kenya Ministry of Health anemia threshold of 100 g/L in early 1999 before deworming treatment. This is low relative to many other areas in Africa, of which many have substantial helminth problems: a recent survey of studies of anemia among school children in less developed countries (Hall and Partnership for*

Child Development (2000)) indicates that there is considerably less anemia in Busia than in samples from Ghana, Malawi, Mali, Mozambique, and Tanzania,"

Aiken *et al.* (2014) downplay the importance of self-reported health, and do not include these measures in their re-analysis. However, self-reported health measures are widely used in studies set in less developed countries, and other research has found that self-reported health often predicts later morbidity and mortality even when other known health risk factors are accounted for (Idler and Benyamini, 1997; Haddock *et al.*, 2006; Brook *et al.*, 1984). In Miguel and Kremer (2004), statistically significant reductions of meaningful magnitude are estimated in both of the self-reported measures ("sick in past week" and "sick often") and both continue to hold in the Aiken et al. (2014) re-analysis.

2.5 Blinding and data quality

Aiken et al. (2014) mention the fact that the Miguel and Kremer (2004) study was unblinded, and make an assertion about the potential effects of this on data quality. On p. 13 they write: *"As this was an unblinded study, Hawthorne effects may have played a role in self-reported outcomes, and fieldworker observations may have been affected by knowledge of intervention status."* They also refer to related points regarding blinding and data quality raised in the recent Cochrane review of deworming (Taylor-Robinson et al., 2012).

It may actually be logistically impossible to carry out a truly unblinded deworming study using a cluster randomized design. Recall that individually randomized designs will be subject to bias in the presence of epidemiological externalities. This makes clustered randomized designs attractive. One of the immediate consequences of taking deworming drugs for many with worm infections is that worms are expelled from the body, usually in stool (although more rarely also through vomiting). This is a highly visible outcome and one that is much commented upon in communities receiving mass deworming.

While individual participants in a study that randomized treatment at the individual level to a subset of children in a school, say, may not know if they received deworming drugs or placebo (since many but not all those who are infected and treated will see worms expelled), participants in a study that randomizes treatment at the cluster level, as in Miguel and Kremer (2004), will immediately know if they are a "treatment" or "placebo" school: in treatment schools, a sizeable group of students (approximately 12% in our data) will immediately experience gastrointestinal discomfort, worms will be expelled in stool and some will vomit; in placebo schools, there will be no such outcomes. Similarly, it would be impossible for enumerators to avoid finding out the school's treatment status, since enumerators interview and speak with hundreds of pupils, teachers and parents during a school visit, and side effects are a common topic of conversation.

Thus a direct, but quite unattractive, implication of Aiken et al. (2014)'s concern with blinding is that it may simply be impossible to carry out a "high quality" double-blinded deworming cluster randomized study. Given these inherent difficulties with "blinding" cluster randomized studies of deworming, unblinded studies like Miguel and Kremer (2004) are likely to constitute the best quality evidence available for making informed judgments about the real-world impacts of deworming. In a context like this, applying standards developed for the far more controlled environment of traditional drug trials may be counterproductive. Instead, given that blinding may be impossible in practice, developing new standards for recognizing good protocols for minimizing bias in data collection in such situations would be useful. As we noted in the original Miguel and Kremer (2004) paper, data collection procedures and timing were balanced across all three groups of schools, and the professional field staff were extensively trained in appropriate and consistent data collection procedures. There are multiple pieces of evidence in the Miguel and Kremer (2004) paper suggesting that data collection was in fact carried out in an even-handed and balanced way across the treatment and control groups, including the fact that pupil school transfer rates, attrition rates, and baseline characteristics are all "balanced" across the three study arms. Moreover, neither of the two key outcome measures – worm infection rates and school attendance – are subjectively measured questions that ask the respondent or the enumerator to make a judgment call. There is simply not much room for unconscious bias to enter into the collection of these variables.

Hawthorne or placebo effects could also not readily explain why there are externality effects on worm infections or on school participation, since these externality effects are realized among pupils in the same school who did not take deworming medicine, or among pupils who were in schools not subject to treatment at all.

The finding of externality effects both within schools and across schools also appears inconsistent with the Aiken et al. team's claim about potentially biased data collection by enumerators (fieldworkers). The study of externalities was not central to the original research design of the study, nor was it an issue that was ever discussed with the field data collection team. It seems inconceivable to us that biased data collection could have generated the results that there are positive worm infection externalities within 0-3 km of treatment schools, and the same holds for the measured health and school participation effects among untreated children within treatment schools.

3. Responses to other points raised in Aiken et al. (2014)

This section provides detailed responses to other points raised in the Aiken et al. (2014) report. For legibility, we have included the original text from the Aiken et al. (2014) report in ***bold italics***, followed by our response. Square brackets denote text added to the quotes for clarity.

Page 1: "Schools were stratified by administrative area and involvement in other ICS programs and then quasi-randomised into three groups..."

This is incorrect. While the original plan had been to stratify by participation in other NGO programs, the actual randomization was not carried out this way. Local schools participating in the intensive CSP/SHP program were dropped from the sample of eligible schools, while 27 primary schools with less intensive NGO programs were retained in the sample. These 27 schools were receiving assistance in the form of either free classroom textbooks, grants for school committees, or teacher training and bonuses. It is worth emphasizing that randomized evaluations of these various interventions did not find statistically significant average project impacts across a wide range of educational outcomes (Glewwe, Kremer, and Moulin, 2009). The schools that benefited from these previous programs were found in all eight geographic zones in the study area. The results in Miguel and Kremer (2004) are robust to including controls for inclusion in these other NGO programs. A complete discussion of the randomization approach is contained in Baird et al. (2014).

Page 2: "In the original analysis, an estimate of the direct benefit of the intervention in each of the three domains [health, school attendance, and exam performance] was obtained by comparing outcomes in schools that received the intervention to those that did not receive it in that year. This format of analysis makes the assumption that there was no cumulative effect of the intervention, such that the second year of the intervention in Group 1 had the same effect as the first year of the intervention in Groups 1 and 2."

We test whether or not effects are the same in the first and second year of deworming treatment in the original paper when the data allows us to do so, and do not find evidence of a differential effect. As we note in Table IX and p. 193 of Miguel and Kremer (2004): *"Point estimates are 6.2 percentage points for the first year of treatment and 4.0 percentage points for the second year, with significance levels of 99 percent and 90 percent, respectively (regression 2), although confidence intervals are wide enough that we cannot reject the hypothesis that the effect is the same in both years."* A similar strategy is used to assess whether there is any differential effect of the first and second years of the intervention on academic test scores in Table X (columns 2 and 3), and we cannot reject the hypothesis of equal impacts in year 1 and year 2. The results presented in Table IX and Table X do this in a way that makes full use of the data, thereby providing maximum statistical power to the analysis.

With regard to the main health outcomes, Miguel and Kremer (2004) estimate differences between treatment and control group individuals in early 1999, after Group 1 individuals had received deworming treatment but before any Group 2 or Group 3 individuals received treatment. Thus we are only able to assess effects after one year of treatment for those outcome measures (in Tables V, VI, and VII).

Note finally that even if the effects differed between years, the main estimated effect in Miguel and Kremer (2004) can be accurately interpreted as the average effect of the program.

Page 2: "As these worm infections are all transmitted by excretion of worm eggs in faeces, and as faecal contamination of the environmental was known to be common, it was assumed that there would be a local reduction of transmission of worm infection in an approximately circular area up to 6 km around the intervention schools, where the children attending the school were assumed to live."

As we describe in detail in the main text of this note, we did not have *a priori* assumptions regarding the precise radius over which there would be epidemiological externalities related to worm infections, nor was there any existing quantitative research to our knowledge that would guide this choice. In fact, we did not even plan to study cross-school externality effects at the start of the study. Thus the focus on a particular radius around each school was guided by analysis of the data. The data suggest that epidemiological and school participation externalities extend out to the 12 closest schools, or about 4 km (see Tables 3 and 4 above), but probably not beyond that.

Page 8: Table 3 (UPDATED Miguel and Kremer, 2004, Table I) results

We disagree with Aiken *et al.* (2014) regarding the means reported for "Grade progression" in the columns for Group 1, Group 2, and Group 3 in this table. We have provided updated values for these figures in Appendix A of this document. We find that no substantive results are changed.

Page 9: "It is not clear from the original report how this randomization [for stool sample testing] was performed."

The random sampling for stool sample testing was conducted nearly seventeen years ago, and unfortunately we are unable to locate the statistical code that produced it. We do know from our recollection of the field work that this was meant to be a representative sample of the students in each grade in the surveyed schools.

Page 9: "Thresholds used for moderate/heavy infection with hookworm, whipworm and schistosomiasis are different from those suggested by the World Health Organization… The supporting reference provided makes a case for and uses locally-defined thresholds for heavy infection (Brooker et al., 2000), but does not mention moderate/heavy infection thresholds. Therefore, a more appropriate description of how these thresholds for moderate/heavy infection were selected would be 'personal communication with Dr Simon Brooker and Dr Donald Bundy', according to the authors' own report of how this was actually done."

We thank the replication authors for clarifying this point. The Brooker *et al.* (2000) article uses alternative thresholds (that do not correspond exactly to the WHO standard) for defining heavy infections in the study area; namely, infection levels at the 90th percentile level. These thresholds are: Hookworms 1,250+ epg; A.lumbricoides 20,000+ epg; T.trichiura 1,000+ epg; S.mansoni 500+ epg, somewhat lower than the WHO standard for heavy infections. Both the heavy infection prevalence used in the Brooker *et al.* (2000) paper and the moderate-to-heavy infection levels used in Miguel and Kremer (2004) were developed in personal communication with Simon Brooker for the specific context of Busia District, Kenya during 1998 and 1999. As these were designed in close consultation with Dr. Brooker and Dr. Donald Bundy, both global experts on intestinal worms, we felt that deviating from the WHO standard was an appropriate adjustment for the setting of the study.

Page 9: "Eggs count data are highly skewed, so presentation of the arithmetic means … is not an appropriate summary of the distribution of values."

There may be differences across disciplines in the presentation of data. The arithmetic mean of egg count data is a meaningful statistic, as are the alternative statistics proposed in Aiken et al. (2014).

Page 13: "The column of results headed 'Group 2', does actually contain some observations based on pupils in Group 3, but as neither Groups did received intervention in year 1 (1998), this has little bearing on the interpretation "

We thank the replication authors for clarifying this point. Although the table notes in Miguel and Kremer (2004) did note that observations from Group 3 were included in the outcomes collected in the 1999 Pupil Questionnaire, it was not clear in the table that for these outcomes, the "Group 2" column actually summarizes data from Groups 2 and 3, and for the "Group 1 – Group 2" column, results are shown for "Group 1 – Groups 2&3". The affected outcome variables include "sick in past week", "height-for-age Z-scores", "weight-for-age Z-scores", "clean" and "days contact with fresh water in past week". We have annotated this table appropriately in the fully updated tables in Appendix A. We agree with Aiken *et al.* (2014)'s interpretation of this as a minor issue, given that both Group 2 and Group 3 were in the control group at the time.

Page 15: ***"Overall, the calculations performed in Table VI represent a set of crude comparisons after just the first year of the study. For all the parasitological outcomes, there is a lack of baseline infection data for Group 2 pupils that means these interpretations may be affected by secular trends in parasite burden – hence these results are not a comparison of "like with like". The original authors were aware of the limitations of these crude comparisons and place no great emphasis on these results in their interpretation, concentrating rather on results of regressions in Tables VII, IX and X that combine data from all Groups across both years of the study."***

We believe it is appropriate to compare Group 1 and Group 2 parasitological outcomes in the absence of baseline data for Group 2. Schools were randomized into intervention groups, and we have shown that baseline balance was achieved across groups along a wide range of health, educational, socioeconomic and other characteristics. Aiken et al. (2014) acknowledge this balance in baseline characteristics on p. 28 of their report. There is no *a priori* reason to believe that parasite burdens were evolving differentially across the three program groups for any reason other than due to the intervention itself. Moreover, the Aiken et al. (2014) report does not provide any reason to believe that there might be differential trends across groups. Indeed, Table VII still focuses only on the results from the first year of the study (using parasitological data that was collected from Group 1 and Group 2 schools in early 1999, prior to Group 2 schools entering into treatment).

We do place the emphasis on Tables VII, IX, and X but that is because these are the tables which use estimation techniques that allow for cross-school externalities, use the full sample of data, and utilize the study's prospective stepped wedge design. This is the most statistically appropriate and powerful approach.

Page 15: "A second coding error was present that miscalculated local density figures for three of the schools – these were School numbers 108 (in Group 1), 109 (in Group 2), and 115 (in Group 3)… This code was problematic as it erroneously assigned these three schools into a "-1" category where their population were ignored when calculating the local densities."

While we appreciate the new results correcting for the error, we should note that this characterization of the second coding error is not correct. For school 108 (a Group 3 school), the coding error resulted in ignoring all Group 1 schools in calculation of the local density terms – however, we note that there were no Group 1 schools located within 6 km of school 108, so the coding error literally had no effect on the data in this case. For school 109 (a Group 2 school), all Group 2 schools were ignored in calculation of the local density terms. There was only 1 Group 2 school located with 6 km of school 109 (and no Group 2 schools located within 3 km of school 109), so this error affected the 3-6 km density term only (by missing one school), not the 0-3 km term for this school. Finally, for school 115 (a

Group 1 school), all Group 3 schools were ignored in calculation of the local density terms, and there were seven such schools within 6 km of school 115. Hence, only two schools were affected by this coding error.

Page 18: Table 10 (UPDATED Miguel and Kremer, 2004, Table VII) results

We disagree with Aiken *et al.* (2014) regarding the results reported in column (6) of this table. In particular, the result on "Group 1 pupils within 3 km (per 1000 pupils)" should read -0.08 (s.e. 0.07). This is not statistically significant at traditional levels of confidence. We have provided an updated version of this table in Appendix A.

Page 20: "[T]he authors weighted by the number of pupil-observations in a school, and not the number of pupils – this is a subtle but important difference."

The large, positive, and statistically significant impacts of deworming on school participation hold whether weighted by the number of pupil-observations or the number of pupils. In fact, both of these approaches are standard in the related research literature and have their merits. Pupil weighting is attractive since it generates the population average, while pupil-observation weights increase power and precision. The school participation results are robust to either approach.

Page 25: "However, the between-school indirect effect [on school attendance] is reversed entirely … [and] is now within one standard error interval of zero, so is clearly non-significant. The overall effect on school attendance is also substantially reduced … and is now only slightly more than one standard error interval away from zero, so is also non-significant."

We address this claim extensively in the main text of this note. The results after correction of variable construction make clear that it is impossible to precisely estimate overall deworming externalities on school participation out to a distance of 6 km, as this results in very wide and largely uninformative confidence intervals, although it is worth noting that the point estimate on the 0-2 km, 0-3 km, and 0-4 km externality terms remains negative, large and statistically significant at 95% confidence (in Table 4, columns 2-4). When we instead explore overall externality effects only up to distances which are precisely estimated, we find large, positive and statistically significant between-school externality impacts (see Tables 2 and 4 above, as well as Figure 2 Panel B).

Page 25: "We note that the original authors did not acknowledge the alteration in these results in their own 2007 internal replication, although the results as found above are shown in their data-recording (log) files from that time."

In response to requests from researchers seeking to use the data or replicate results, we assembled detailed documentation, data files (in STATA), statistical analysis files (do files), and results files (log files) in 2007, and freely shared these materials with researchers and non-profit organizations interested in re-analyzing the data. These materials contained updated data that corrected the coding error in the construction of the 3-6 km externality term. Between 2007 and 2013, we shared these detailed and updated replication materials with researchers at Columbia University, Georgia State University, Givewell, Hampshire College, Harvard University, the MIT Abdul Latif Jameel Poverty Action Lab, the University of California, Berkeley, the University of Glasgow, Vanderbilt University, and Yale University.

(We believe there were likely other researchers with whom we shared the data but did not keep track of, so this is a lower bound on the total number). They represented a range of fields, including economics, political science, development studies, education policy, health economics, and public policy. The materials included the "log" files where statistical results are reported in STATA.

In many economics Ph.D. programs, graduate students are asked to replicate a well-known empirical paper in an econometrics or statistical methods class; in cases where substantial discrepancies are found, these replication exercises are then often published in peer-reviewed economics journals as a comment on the original paper. Several of the faculty and graduate student researchers who requested and received the replication materials mentioned that they were in fact seeking the Miguel and Kremer (2004) materials for such a course assignment. It is worth noting that none of these scholars considered the differences with the original paper sufficiently important to publish any of their findings.

Page 25: "This within-school indirect effect [on school attendance] in similar to the direct effect of the intervention… but note that both of these results represent data from the first year of the study only. This prompts the question – if all children in treatment schools had fairly similar benefit on school attendance, regardless of whether or not they received the drug treatment for de-worming, then how much of the effect on school attendance is attributable to drug administration?"

As discussed in Miguel and Kremer (2004), there is strong evidence of deworming externalities on both worm infections and school participation both within the treatment schools (the focus of the quote above) and to neighboring schools within 3 km. There are large school participation gains for both the treated and untreated pupils in treatment schools, and we cannot reject that the impacts are equal for these two groups (in Table IX of Miguel and Kremer 2004, and in Aiken et al. 2014), in part because this is a relatively statistically underpowered test. Deworming breaks the cycle of transmission for worm infections, reducing reinfection for individuals within and in the vicinity of treatment schools. Intestinal worms have quite short average lifespans, on the order of one to two years, so sharp reductions in reinfection could quickly translate into a lower worm disease burden among both the untreated and the treated. Other work also suggests substantial epidemiological externalities among the untreated in treatment communities (Bundy et al. 1990, Ozier 2014).

This question posed by Aiken et al. (2014) also raises the issue of whether there are other factors beyond reduced worm infections driving impacts on school participation. In other parts of their report, for instance, they raise the possibility of Hawthorne-type effects. Hawthorne effects and placebo seem highly unlikely to be a concern for the untreated children in treatment schools; since they did not receive medical treatment themselves, it is difficult to argue that they would benefit from a placebo effect. There is also a related possibility, namely that it was health education rather than the deworming drugs that drove impacts. However, we show in both Miguel and Kremer (2004) and Kremer and Miguel (2007) that there are no significant differences in a range of worm prevention behaviors between the treatment and control schools, including wearing shoes, contact with fresh water, or observed cleanliness. Given the implausibility of placebo effects for those not receiving treatment, and the lack of evidence of any sort of positive health behavioral change, the leading explanation is that worm infection externality benefits are driving the school participation externalities (as we argued in Miguel and Kremer, 2004).

Page 26: Table 16 (UPDATED Miguel and Kremer, 2004, Table X) results

We disagree with Aiken *et al.* (2014) regarding the results reported in column (2) of this table of their report. In particular, the standard error on "Second year as treatment school (T2)" is 0.079. We have provided an updated version of this table in Appendix A. No substantive results are changed.

Page 28: "…the spread of [school] sizes is indeed broadly similar between the groups, but there are some extreme sizes in Group 2."

This outlier in school population among Group 2 schools is because two study sample schools (both Group 2 schools) were flooded in late 1997, and were not open during the 1998 school year. Most of these students ended up enrolling in another nearby PSDP school, which was a Group 2 school, temporarily swelling its enrolment. Most of the pupils returned to their original schools in 1999. Note that we followed a standard intention to treat (ITT) approach and continue to assign each pupil to her/his original school throughout the study.

Page 29: "***The extent of missing data varies enormously in this study. Within Table I, for example, in the 'Year of Birth' variable in Panel A, 22% of values for individual children are missing, but in the Weight-for-Age z-score variable in panel B less than 0.1% of data are missing. This is perplexing as age was presumably known for all children in whom Weight-for-Age was calculated – so why was this age data not used to complete missing age values elsewhere?"***

The year of birth sample that the authors reference includes all 34,792 children in the "namelist" data set for the program, while the WAZ figure is only available for the subset of children who were administered the 1998 Pupil Questionnaire, a group of 13,130 of the namelist children, for all but 3 of whom age data was not missing in the "namelist" file. The questionnaire was only administered in grades 3 through 8, and not in grades 1 and 2 (nor in the kindergarten/nursery class, grade 0). The year of birth data was much more likely to be missing for children in grades 0, 1, and 2 (who were not administered the 1998 Pupil Questionnaire). It is not unexpected for many children in this young age group in the study setting to be unaware of their exact birth year or birthdate.

Page 29: "All the tables in the study make use of a standard approach of annotating the results significant at 99 (***), 95 (**) and 90 (*) percent confidence, rather than displaying actual p-values. Whilst there is certainly merit in not restricting scientific investigation to using a significance level of 95% (or α of 0.05), it should be borne in mind that results at 90% confidence could have occurred by chance on up to 1 in 10 occasions."

This is a standard approach of annotating statistical significance in economics and many other fields.

Page 29: "On a wider level, there is the question of what would have been the most appropriate level of significance to use in these analyses where multiple comparisons were being performed. Given that there was not a single pre-

specified question being studied, but rather multiple exploratory comparisons across many different categories of outcome, it would seem prudent to have considered use of a Bonferroni correction for multiple testing…"

This is an interesting and potentially important point, not just for Miguel and Kremer (2004) but for many if not most empirical papers in economics, public health, and other research fields. To our knowledge, there is no standard statistical approach for addressing multiple inference concerns in analyses that were not pre-specified. (In contrast, in pre-specified analyses, it is straightforward to determine the universe of statistical tests that were performed and to adjust p-values accordingly.)

The standard approach in economics is to test hypotheses in multiple ways and assess if there is a consistent pattern of results. Indeed, four pieces of evidence point towards epidemiological externalities, namely, both the within-school and the cross-school externalities within 0-3 km, for both worm infections and school participation. Recently, the Ozier (2014) study provides a fifth piece of evidence for positive deworming treatment spillovers in these communities (in his case, among those who were infants at the time deworming took place).

Page 29: Given that in African countries, many (or all) of these factors are highly variable in both time and place, it is certainly possible that the indirect between-school effects could be substantially greater or smaller in a different location. This would mean that the 'overall' result as reported in this study might be less generalizable than the 'naïve' effect as the latter does not try to take account of the between-school indirect effect that might have such wide geographical variation."

Both "direct" effects on the treated and indirect effects on the untreated are certainly a function of local circumstances. For example, direct effects could vary with the extent of worm prevalence, with malaria and other health factors, and with incentives and social norms regarding school attendance. Externalities could vary with population density, with patterns of social interaction in the community, etc. We do not see a prior reasons why the naïve effect, which we show will be statistically biased in the presence of externalities, will be more generalizable.

Page 30: "In contrast to the original report, our pure replication found little evidence of non-worm-related health benefits…"

This conclusion is as much due to a focus by the replication authors on particular health measures as to a change in the results of Miguel and Kremer (2004). The only variable with significance changes was the "proportion anemic", where the coefficient estimate is unchanged but where the P-value was reported as being less than 0.05 when it is actually 0.19. The statistical significance of the level of Hb, the self-reported health outcomes, and the HAZ (height) result are unchanged in the replication results presented in Aiken et al. (2014).

References

Aiken AM, Davey C, Hayes RJ, Hargreaves J. Deworming schoolchildren in Kenya - Replication plan. International Institute Impact Evaluation (3ie) website: 2013.

Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ. (2014). "Reanalysis of health and educational impacts of a school-based deworming program in western Kenya: Part 1, pure replication", 3ie Replication Paper 3, part 1. Washington, DC: International Initiative for Impact Evaluation (3ie).

Alderman, H., J. Konde-Lule, I. Sebuliba, D. Bundy, A. Hall. (2006). "Increased weight gain in preschool children due to mass albendazole treatment given during 'Child Health Days' in Uganda: A cluster randomized controlled trial", *British Medical Journal*, 333, 122-6.

Alderman, Harold. (2007). "Improving nutrition through community growth promotion: Longitudinal study of nutrition and early child development program in Uganda", *World Development*, 35(8), 1376-1389.

Awasthi, Shally, et al. (2013). "Population deworming every 6 months with albendazole in 1 million pre-school children in north India: DEVTA, a cluster-randomized trial", Lancet, 381(9876): 1478-1486.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. (2014). "Worms at Work: Public finance implications of a child health investment", unpublished working paper, University of California, Berkeley.

Brook, R.H., et al. (1984). *The Effect of Coinsurance on the Health of Adults: Results from the RAND Health Insurance Experiment*. RAND: Santa Monica, CA.

Bundy, D.A.P., M.S. Wong, L.L. Lewis, and J. Horton. (1990). "Control of Geohelminths by Delivery of Targeted Chemotherapy through Schools", *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 84, 115-120.

Conley, Timothy G., and Christopher R. Udry. (2010). "Learning about the New Technology: Pineapple in Ghana", *American Economic Review*, 100(1): 35-69.

Croke, Kevin. (2014). "The long run effects of early childhood deworming on literacy and numeracy: Evidence from Uganda", unpublished working paper, Harvard University.

Haddock, C.K., et al. (2006). "The validity of self-rated health as a measure of health status among young military personnel: evidence from a cross-sectional survey", *Health and Quality of Life Outcomes*, 4(57).

Idler, Ellen L., and Yael Benyamini. (1997). "Self-rated health and mortality: A review of twenty-seven community studies", *Journal of Health and Social Behavior*, 38(1).

Kremer, Michael, and Edward Miguel. (2007). "The Illusion of Sustainability", *Quarterly Journal of Economics*, 112(3), 1007-1065.

Miguel, Edward and Michel Kremer (2001). "Worms: Education and Health Externalities in Kenya", National Bureau of Economic Research Working Paper #8481.

Miguel, Edward and Michael Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72(1), 159-217.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn, M. Van der Laan. (2014). "Promoting Transparency in Social Science Research", *Science*, 10.1126/science.1245317.

Ozier, Owen. (2014). “Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming”, World Bank Policy Research Working Paper #7052.

Taylor-Robinson DC, Maayan, N, Soares-Weiser, K. Garner, P. (2012) “Deworming drugs for treating soil-transmitted intestinal worms in children: effects on nutrition and school performance.” *Cochrane Database of Systematic Reviews*.

Appendix A: Updated tables for Miguel and Kremer (2004)

This appendix includes all tables in Miguel and Kremer (2004), updated to use the “final” versions of all datasets and corrected of all rounding, typographical and coding errors.

Table I: 1998 Average pupil and school characteristics, pre-treatment

	Group 1 (25 schs)	Group 2 (25 schs)	Group 3 (25 schs)	G1–G3	G2–G3
Panel A: Pre-school to Grade 8					
Male	0.53	0.51	0.52	0.01	-0.01
				(0.02)	(0.02)
Proportion girls < 13 years, and all boys	0.89	0.89	0.88	0.00	0.01
				(0.01)	(0.01)
Grade progression (= Grade–(Age–6))	-2.0	-1.8	-2.0	-0.0	0.1
				(0.1)	(0.1)
Year of birth	1986.2	1986.5	1985.8	0.4**	0.8***
				(0.2)	(0.2)
Panel B: Grades 3 to 8					
Attendance recorded in school registers (during 4 weeks prior to pupil survey)	0.973	0.963	0.969	0.003	-0.006
				(0.004)	(0.004)
Access to latrine at home	0.82	0.81	0.82	0.00	-0.01
				(0.03)	(0.03)
Have livestock (cows, goats, pigs, sheep) at home	0.66	0.67	0.66	-0.00	0.01
				(0.03)	(0.03)
Weight-for-age Z-score (low scores denote undernutrition)	-1.39	-1.40	-1.44	0.05	0.04
				(0.05)	(0.05)
Blood in stool (self-reported)	0.26	0.22	0.19	0.07**	0.03
				(0.03)	(0.03)
Sick often (self-reported)	0.10	0.10	0.08	0.02	0.02*
				(0.01)	(0.01)
Malaria/fever in past week (self-reported)	0.37	0.38	0.40	-0.03	-0.02
				(0.03)	(0.03)
Clean (observed by field workers)	0.60	0.66	0.67	-0.07**	-0.01
				(0.03)	(0.03)
Panel C: School characteristics					
District exam score 1996, grades 5-8‡	-0.10	0.09	0.01	-0.11	0.08
				(0.12)	(0.12)
Distance to Lake Victoria	10.0	9.9	9.5	0.6	0.5
				(1.9)	(1.9)
Pupil population	392.7	403.8	375.9	16.8	27.9
				(57.6)	(57.6)
School latrines per pupil	0.007	0.006	0.007	0.001	-0.000
				(0.001)	(0.001)
Proportion moderate-heavy infections in zone	0.37	0.37	0.36	0.01	0.01
				(0.03)	(0.03)
Group 1 pupils within 3 km††	430.4	433.2	344.5	85.9	88.7
				(116.2)	(116.2)
Group 1 pupils within 3-6 km	1157.6	1043.0	1297.3	-139.7	-254.4
				(199.3)	(199.3)
Total primary school pupils within 3 km	1272.7	1369.1	1151.9	120.8	217.2
				(208.1)	(208.1)
Total primary school pupils within 3-6 km	3431.3	3259.8	3502.1	-70.8	-242.3
				(366.0)	(366.0)

Note: School averages weighted by pupil population. Standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. Data from the 1998 ICS Pupil Namelist, 1998 Pupil Questionnaire and 1998 School Questionnaire. ‡1996 District exam scores have been normalized to be in units of individual level standard deviations, and so are comparable in units to the 1998 and 1999 ICS test scores (under the assumption that the decomposition of test score variance within and between schools was the same in 1996, 1998, and 1999). †† This includes girls less than 13 years old, and all boys (those eligible for deworming in treatment schools).

Table II: January 1998 helminth infections, pre-treatment, Group 1 schools

	Prevalence of infection	Prevalence of moderate-heavy infection	Average worm load, in eggs per gram (s.e.)
Hookworm	0.77	0.15	426 (1055)
Roundworm	0.42	0.16	2337 (5156)
Schistosomiasis, all schools	0.22	0.07	91 (413)
Schistosomiasis, schools < 5km from Lake Victoria	0.80	0.39	487 (879)
Whipworm	0.55	0.10	161 (470)
At least one infection	0.92	0.37	-
Born since 1985	0.93	0.40	-
Born before 1985	0.91	0.34	-
Female	0.91	0.34	-
Male	0.93	0.38	-
At least two infections	0.65	0.10	-
At least three infections	0.34	0.01	-

Note: These are averages of individual-level data, as presented in Brooker, et al. (2000b); correcting for the oversampling of the (numerically smaller) upper grades does not substantially change the results. Standard errors in parentheses. Sample size: 1894 pupils. Fifteen pupils per standard in grades 3 to 8 for Group 1 schools were randomly sampled. The bottom two rows of the column "Prevalence of moderate-heavy infection" should be interpreted as the proportion with at least two or at least three moderate-to-heavy helminth infections, respectively. The data were collected in January to March 1998 by the Kenya Ministry of Health, Division of Vector Borne Diseases (DVBD). The moderate infection thresholds for the various intestinal helminths are: 250 epg for *S. mansoni*, and 5,000 epg for Roundworm, both the WHO standard, and 750 epg for Hookworm and 400 epg for Whipworm, both somewhat lower than the WHO standard. Refer to Brooker, et al. (2000b) for a discussion of this parasitological survey and the infection cut-offs. All cases of schistosomiasis are *S. mansoni*.

Table III: Proportion of pupils receiving deworming treatment in PSDP

	Group 1		Group 2		Group 3	
	Girls < 13 yrs, all boys	Girls ≥ 13 yrs	Girls < 13 yrs, all boys	Girls ≥ 13 yrs	Girls < 13 yrs, all boys	Girls ≥ 13 yrs
	Treatment		*Comparison*		*Comparison*	
Any medical treatment in 1998 (For grades 1-8 in early 1998)	0.77	0.20	0	0	0	0
Round 1 (March-April 1998), Albendazole	0.68	0.11	0	0	0	0
Round 1 (March-April 1998), Praziquantel‡	0.64	0.34	0	0	0	0
Round 2 (Oct.-Nov. 1998), Albendazole	0.56	0.07	0	0	0	0
	Treatment		*Treatment*		*Comparison*	
Any medical treatment in 1999 (For grades 1-7 in early 1998)	0.58	0.07	0.54	0.09	0.01	0
Round 1 (March-June 1999), Albendazole	0.44	0.06	0.35	0.05	0.01	0
Round 1 (March-June 1999), Praziquantel‡	0.47	0.06	0.38	0.06	0.00	0
Round 2 (Oct.-Nov. 1999), Albendazole	0.52	0.06	0.50	0.07	0.01	0
Any medical treatment in 1999 (For grades 1-7 in early 1998), among pupils enrolled in 1999	0.73	0.10	0.71	0.14	0.02	0
Round 1 (March-June 1999), Albendazole	0.55	0.08	0.46	0.08	0.01	0
Round 1 (March-June 1999), Praziquantel‡	0.54	0.08	0.46	0.07	0.00	0
Round 2 (Oct.-Nov. 1999), Albendazole	0.65	0.09	0.66	0.11	0.01	0

Note: Data for grades 1-8. Since month of birth information is missing for most pupils, precise assignment of treatment eligibility status for girls born during the "threshold" year is often impossible; all girls who turn 13 during a given year are counted as 12 year olds (eligible for deworming treatment) throughout for consistency.

‡Praziquantel figures in Table 3 refer only to children in schools meeting the schistosomiasis treament threshold (30 percent prevalence) in that year.

Table IV: Proportion of pupil transfers across schools

	1998 transfer to a			1999 transfer to a		
School in early 1998 (pre-treatment)	Group 1 School	Group 2 School	Group 3 school	Group 1 school	Group 2 school	Group 3 school
Group 1	0.005	0.007	0.007	0.032	0.026	0.027
Group 2	0.006	0.007	0.008	0.026	0.033	0.027
Group 3	0.010	0.010	0.006	0.022	0.036	0.022
Total transfers	0.020	0.024	0.020	0.080	0.095	0.076

Table V: January to March 1999, Health and Health Behavior Differences Between Group 1 (1998 Treatment) and Group 2 (1998 Comparison) Schools

	Group 1	Group 2°	G1 – G2°
Panel A: Helminth Infection Rates			
Any moderate-heavy infection, January – March 1998	0.38	-	-
Any moderate-heavy infection, 1999	0.27	0.52	-0.25*** (0.06)
Hookworm moderate-heavy infection, 1999	0.06	0.22	-0.16*** (0.03)
Roundworm moderate-heavy infection, 1999	0.09	0.24	-0.15*** (0.04)
Schistosomiasis moderate-heavy infection, 1999	0.08	0.18	-0.10 (0.06)
Whipworm moderate-heavy infection, 1999	0.13	0.17	-0.04 (0.05)
Panel B: Other Nutritional and Health Outcomes			
Sick in past week (self-reported), 1999	0.40	0.45	-0.05** (0.02)
Sick often (self-reported), 1999	0.12	0.15	-0.03** (0.01)
Height-for-age Z-score, 1999 (low scores denote undernutrition)	-1.13	-1.22	0.08* (0.05)
Weight-for-age Z-score, 1999 (low scores denote undernutrition)	-1.25	-1.25	-0.00 (0.04)
Hemoglobin concentration (g/L), 1999	124.9	123.3	1.6 (1.4)
Proportion anemic (Hb < 100g/L), 1999	0.02	0.04	-0.02 (0.01)
Panel C: Worm Prevention Behaviors			
Clean (observed by field worker), 1999	0.59	0.60	-0.01 (0.02)
Wears shoes (observed by field worker), 1999	0.24	0.26	-0.02 (0.03)
Days contact with fresh water in past week (self-reported), 1999	2.4	2.2	0.2 (0.3)

Note: These are averages of individual-level data for grade 3-8 pupils; disturbance terms are clustered within schools. Robust standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. Obs. for parasitological results: 2328 (862 Group 1,466 Group 2). Obs. for hemoglobin results: 769 (290 Group 1, 479 Group 2). Obs. for 1999 Pupil Questionnaire health outcomes: 9,039 (3545 Group 1, 5497 Group 2 and Group 3). Following Brooker et al. (2000b), moderate-to-heavy infection thresholds for the various intestinal helminths are: 250 epg for *S. mansoni*, and 5,000 epg for Roundworm, both the WHO standard, and 750 epg for Hookworm and 400 epg for Whipworm, both somewhat lower than the WHO standard. Kenya Ministry of Health officials collected the parasitological data from January to March 1998 in Group 1 schools, and from January to March 1999 in Group 1 and Group 2 schools. A random subset of the original 1998 Group 1 parasitological sample was re-surveyed in 1999. Hb data were collected by Kenya Ministry of Health officials and ICS field officers using the portable Hemocue machine. The self-reported health outcomes were collected for all three groups of schools as part of Pupil Questionnaire administration.

° Note that for the outcomes collected in the 1999 Pupil Questionnaire, statistics in these columns also include Group 3 individuals.

Table VI: Deworming health externalities within schools, January to March 1999

	G1, Treated in 1998	G1, Untreated in 1998	G2, Treated in 1999	G2, Untreated in 1999	(G1 Treated 1998) – (G2, Treated 1999)	(G1, Untreated 1998) – (G2, Untreated 1999)
<u>Panel A:</u> *Selection into Treatment*						
Any moderate-heavy infection, 1998	0.39	0.44	-	-	-	-
Proportion of 1998 parasitological sample tracked to 1999 sample‡	0.36	0.35	-	-	-	-
Access to latrine at home, 1998	0.85	0.80	0.81	0.86	0.03 (0.04)	-0.06 (0.05)
Grade progression (=Grade – (Age – 6)), 1998	-2.0	-1.8	-1.8	-1.8	-0.2 (0.1)	-0.0 (0.2)
Weight-for-age (Z-score), 1998 (low scores denote undernutrition)	-1.58	-1.52	-1.57	-1.46	-0.01 (0.06)	-0.06 (0.11)
Malaria/fever in past week (self-reported), 1998	0.37	0.41	0.40	0.39	-0.03 (0.04)	0.02 (0.06)
Clean (observed by field worker), 1998	0.53	0.59	0.60	0.66	-0.07 (0.05)	-0.07 (0.10)
<u>Panel B:</u> *Health Outcomes*						
Girls < 13 years, and all boys						
Any moderate-heavy infection, 1999	0.24	0.34	0.51	0.55	-0.27*** (0.06)	-0.21** (0.10)
Hookworm moderate-heavy infection, 1999	0.04	0.11	0.22	0.20	-0.19*** (0.03)	-0.10* (0.05)
Roundworm moderate-heavy infection, 1999	0.08	0.12	0.22	0.30	-0.14*** (0.04)	-0.18** (0.07)
Schistosomiasis moderate-heavy infection, 1999	0.09	0.08	0.20	0.13	-0.11* (0.06)	-0.05 (0.06)
Whipworm moderate-heavy infection, 1999	0.12	0.16	0.16	0.20	-0.04 (0.05)	-0.05 (0.09)
Girls ≥ 13 years						
Any moderate-heavy infection, 1998	0.31	0.30	-	-	-	-
Any moderate-heavy infection, 1999	0.27	0.44	0.32	0.54	-0.05 (0.17)	-0.09 (0.09)
<u>Panel C:</u> *School Participation*						
School participation rate, May 1998 to March 1999††	0.872	0.774	0.808	0.690	0.064* (0.033)	0.084** (0.037)

Note: These are averages of individual-level data for grade 3-8 pupils in the parasitological survey subsample; disturbance terms are clustered within schools. Robust standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. The data are described in the footnote to Table 5. Obs. for the 1999 parasitological survey: 669 Group 1 treated 1998, 76 Group 1 untreated 1998, 874 Group 2 treated 1999, 349 Group 2 untreated 1999. ‡We attempted to track a random sample of half of the original 1998 parasitological sample. Because some pupils were absent, had dropped out, or had graduated, we were only able to re-survey 72 percent of this subsample. ††School averages weighted by pupil population. The participation rate is computed among pupils enrolled in the school at the start of 1998. Pupils present in school during an unannounced NGO visit are considered participants. Pupils had 3.8 participation observations per year on average. Participation rates are for grades 1 to 7; grade 8 pupils are excluded since many graduated after the 1998 school year, in which case their 1999 treatment status is irrelevant. Preschool pupils are excluded since they typically have missing compliance data. All 1998 pupil characteristics in Panel A are for grades 3 to 7, since younger pupils were not administered the Pupil Questionnaire.

Table VII: Deworming health externalities within and across schools, January to March 1999

	Any moderate-heavy helminth infection, 1999			Moderate-heavy schistosomiasis infection, 1999			Moderate-heavy geohelminth infection, 1999		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Indicator for Group 1 (1998 Treatment) School	-0.31***	-0.18**	-0.21*	-0.09***	-0.06	-0.03	-0.30***	-0.19***	-0.26***
	(0.06)	(0.07)	(0.11)	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)	(0.09)
Group 1 pupils within 3 km (per 1000 pupils)	-0.21**	-0.22**	-0.10	-0.12***	-0.12***	-0.08	-0.12	-0.13	-0.06
	(0.10)	(0.11)	(0.14)	(0.05)	(0.05)	(0.07)	(0.09)	(0.10)	(0.12)
Group 1 pupils within 3-6 km (per 1000 pupils)	-0.05	-0.04	-0.08	-0.15***	-0.15***	-0.13**	0.06	0.08	0.03
	(0.08)	(0.08)	(0.11)	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)	(0.09)
Total pupils within 3 km (per 1000 pupils)	0.05	0.05	0.05	0.08***	0.08***	0.08***	-0.01	-0.01	-0.01
	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Total pupils within 3-6 km (per 1000 pupils)	-0.02	-0.03	-0.02	0.04*	0.04*	0.04*	-0.04	-0.05	-0.04
	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)		-0.06*			0.04**			-0.10***	
		(0.03)			(0.02)			(0.03)	
(Group 1 Indicator) * Received treatment, when offered		-0.15**			-0.04			-0.11**	
		(0.06)			(0.04)			(0.05)	
(Group 1 Indicator) * Group 1 pupils within 3 km (per 1000 pupils)			-0.27**			-0.07			-0.16
			(0.14)			(0.08)			(0.11)
(Group 1 Indicator) * Group 1 pupils within 3-6 km (per 1000 pupils)			0.01			-0.03			0.03
			(0.09)			(0.06)			(0.07)
Grade indicators, controls for school assistance, district exam score	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	2330	2329	2330	2330	2329	2330	2330	2329	2330
Mean of dependent variable	0.41	0.41	0.41	0.16	0.16	0.16	0.32	0.32	0.32

Note: Grade 3-8 pupils. Probit estimation, robust standard errors in parentheses. Disturbance terms are clustered within schools. Observations are weighted by total school population. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. The 1999 parasitological survey data are for Group 1 and Group 2 schools. The pupil population data is from the 1998 School Questionnaire. The geohelminths are hookworm, roundworm, and whipworm. We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools.

Table VIII: School participation, school-level data

	Group 1 (25 schools)	Group 2 (25 schools)	Group 3 (25 schools)		
Panel A: First year post-treatment (May 1998 to March 1999)	*1st Year Treatment*	*Control*	*Control*	*G1–(G2&3)*	*G2 – G3*
Girls < 13 years, and all boys	0.841	0.731	0.766	0.093*** (0.030)	-0.035 (0.035)
Girls ≥ 13 years	0.868	0.804	0.820	0.056* (0.031)	-0.016 (0.036)
Preschool, Grade 1, Grade 2 in early 1998	0.797	0.689	0.707	0.100*** (0.037)	-0.019 (0.043)
Grade 3, Grade 4, Grade 5 in early 1998	0.877	0.788	0.827	0.071*** (0.024)	-0.039 (0.029)
Grade 6, Grade 7, Grade 8 in early 1998	0.934	0.859	0.891	0.058*** (0.021)	-0.032 (0.025)
Recorded as "dropped out" in early 1998	0.066	0.051	0.030	0.024 (0.018)	0.022 (0.017)
Females‡	0.855	0.771	0.789	0.076*** (0.027)	-0.018 (0.032)
Males	0.844	0.736	0.780	0.088*** (0.031)	-0.044 (0.037)
Panel B: Second year post-treatment (March to November 1999)	*2nd Year Treatment*	*1st Year Treatment*	*Control*	*G1 – G3*	*G2 – G3*
Girls < 13 years, and all boys	0.716	0.718	0.664	0.051* (0.027)	0.054* (0.027)
Girls ≥ 14 years††	0.627	0.649	0.588	0.039 (0.035)	0.061* (0.035)
Preschool, Grade 1, Grade 2 in early 1998	0.692	0.725	0.641	0.051 (0.034)	0.084** (0.034)
Grade 3, Grade 4, Grade 5 in early 1998	0.749	0.766	0.720	0.029 (0.022)	0.046** (0.023)
Grade 6, Grade 7, Grade 8 in early 1998	0.781	0.790	0.754	0.027 (0.025)	0.036 (0.026)
Recorded as "dropped out" in early 1998	0.188	0.130	0.062	0.126* (0.066)	0.068 (0.056)
Females‡	0.716	0.746	0.649	0.067** (0.027)	0.097*** (0.027)
Males	0.698	0.695	0.655	0.043 (0.028)	0.040 (0.029)

Note: The results are school averages weighted by pupil population. Standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. The participation rate is computed among all pupils enrolled in the school at the start of 1998. Pupils who are present in school on the day of an unannounced NGO visit are considered participants. Pupils had 3.8 participation observations per year on average. The figures for the "Preschool-Grade 2"; "Grade 3-5"; "Grade 6-8"; and "Dropout" rows are for girls < 13 years, and all boys.

‡396 pupils in the sample are missing information on gender. For this reason, the average of the female and male participation rates does not equal the overall average.

††Examining girls ≥14 years old eliminates the cohort of girls in Group 1 schools (12 year olds in 1998) who were supposed to receive deworming treatment in 1998.

Table IX: School participation, direct effects and externalities
Dependent variable: Average individual school participation, by year

	OLS (1)	OLS (2)	OLS (3)	OLS (4) May 98-March 99	OLS (5) May 98-March 99	OLS (6) May 98-March 99	IV-2SLS (7) May 98-March 99
Moderate-heavy infection, early 1999						-0.025** (0.010)	-0.195** (0.096)
Treatment school (T)	0.057*** (0.014)						
First year as treatment school (T1)		0.063*** (0.015)	0.062*** (0.014)	0.062*** (0.022)	0.056*** (0.020)		
Second year as treatment school (T2)		0.039* (0.021)	0.033 (0.021)				
Treatment school pupils within 3 km (per 1000 pupils)			0.040* (0.022)		0.022 (0.032)		
Treatment school pupils within 3-6 km (per 1000 pupils)			-0.024 (0.015)		-0.067*** (0.020)		
Total pupils within 3 km (per 1000 pupils)			-0.031** (0.012)		-0.040** (0.016)	0.014 (0.014)	-0.029* (0.016)
Total pupils within 3-6 km (per 1000 pupils)			0.012 (0.009)		0.035*** (0.011)	0.016* (0.009)	0.008 (0.009)
Indicator received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)					0.104*** (0.014)		
(First year as treatment school Indicator)* (Received treatment, when offered)					-0.013 (0.020)		
1996 district exam score, school average	0.071*** (0.021)	0.070*** (0.021)	0.077*** (0.022)	0.058* (0.032)	0.106*** (0.034)	0.020 (0.024)	-0.000 (0.022)
Grade indicators, school assistance controls, and time controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.22	0.22	0.22	0.33	0.37	0.29	-
Root MSE	0.279	0.279	0.278	0.223	0.217	0.150	0.069
Number of observations	56496	56496	56496	18215	18215	2327	49 (schools)
Mean of dependent variable	0.747	0.747	0.747	0.793	0.793	0.884	0.884

Note: The dependent variable is average individual school participation in each year of the program (Year 1 is to March 1999, and Year 2 is May 1999 to November 1999); disturbance terms are clustered within schools. Robust standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. Additional explanatory variables include an indicator variable for girls < 13 years and all boys, and the rate of moderate-heavy infections in geographic zone, by grade (zonal infection rates among grade 3 and 4 pupils are used for pupils in grades 4 and below and for pupils initially recorded as drop-outs as there is no parasitological data for pupils below grade 3; zonal infection rates among grade 5 and 6 pupils are used for pupils in grades 5 and 6, and similarly for grades 7 and 8). Participation is computed among all pupils enrolled at the start of the 1998 school year. Pupils present during an unannounced NGO school visit are considered participants. Pupils had approximately 3.8 attendance observations per year. Regressions 6 and 7 include pupils with parasitological information from early 1999, restricting the sample to a random subset of Group 1 and Group 2 pupils. The number of treatment school pupils from May 1998 to March 1999 is the number of Group 1 pupils, and the number of treatment school pupils after March 1999 is the number of Group 1 and Group 2 pupils. The instrumental variables in regression 7 are the Group 1 (treatment) indicator variable, Treatment school pupils within 3 km, Treatment school pupils within 3-6 km, and the remaining explanatory variables. We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools.

Table X: Academic examinations, individual-level data

	Dependent variable: ICS Exam Score (normalized by standard) (1)	(2)	(3) Among those who filled in the 1998 pupil survey
Average school participation (during the year of the exam)	0.63***		
	(0.07)		
First year as treatment school (T1)		-0.035	-0.036
		(0.047)	(0.049)
Second year as treatment school (T2)		-0.015	-0.013
		(0.079)	(0.088)
1996 District exam score, school average	0.74***	0.72***	0.75***
	(0.07)	(0.07)	(0.07)
Grade indicators, school assistance controls, and local pupil density controls	Yes	Yes	Yes
R^2	0.14	0.13	0.15
Root MSE	0.919	0.923	0.916
Number of observations	24979	24979	19072
Mean of dependent variable	0.019	0.019	0.039

Note: Each data point is the individual-level exam result in a given year of the program (either 1998, or 1999); disturbance terms are clustered within schools. Linear regression, robust standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. Regression 3 includes only pupils who completed the 1998 Pupil Questionnaire. Additional explanatory variables include an indicator variable for girls < 13 years and all boys, and the rate of moderate-to-heavy infections in geographic zone, by grade (zonal infection rates among grade 3 and 4 pupils are used for pupils in grades 4 and below and for pupils initially recorded as dropouts as there is no parasitological data for pupils below grade 3; zonal infection rates among grade 5 and 6 pupils are used for pupils in grades 5 and 6, and similarly for grades 7 and 8). The local pupil density terms include treatment school pupils within 3 km (per 1000 pupils), total pupils within 3 km (per 1000 pupils), treatment school pupils within 3-6 km (per 1000 pupils), and total pupils within 3-6 km (per 1000 pupils). We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools. The ICS tests for 1998 and 1999 were similar in content, but differed in two important respects. First, the 1998 exam featured multiple-choice questions while the 1999 test featured short answers. Second, while each grade in 1998 was administered a different exam, in 1999 the same exam – featuring questions across a range of difficulty levels – was administered to all pupils in grades 3 to 8. Government district exams in English, Maths, Science-Agriculture, Kiswahili, Geography-History, Home Science, and Arts-Crafts were also administered in both years. Treatment effect estimates are similar for both sets of exams (results not shown).

Table A2: Local densities of other primary schools and deworming compliance rates

	Dependent variable: 1998 Compliance rate (any medical treatment) OLS (1)	1999 Compliance rate (any medical treatment) OLS (2)
Treatment school pupils within 3 km (per 1000 pupils)	-0.04	-0.04
	(0.07)	(0.12)
Treatment school pupils within 3-6 km (per 1000 pupils)	0.08	-0.01
	(0.05)	(0.06)
Total pupils within 3 km (per 1000 pupils)	0.09**	0.04
	(0.03)	(0.08)
Total pupils within 3-6 km (per 1000 pupils)	-0.03	0.00
	(0.03)	(0.03)
Grade indicators, school assistance controls, district exam score control	Yes	Yes
R^2	0.67	0.56
Root MSE	0.075	0.133
Number of observations	25	49
Mean of dependent variable	0.76	0.51

Note: Robust standard errors in parentheses. Observations are weighted by total school population. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. The 1998 compliance data is for Group 1 schools, and the 1999 compliance data is for Group 1 and Group 2 schools. The pupil population data is from the 1998 School Questionnaire. We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools. The number of treatment school pupils in 1998 is the number of Group 1 pupils, and the number of treatment school pupils in March 1999 is the number of Group 1 and Group 2 pupils.

Table A3: Deworming health externalities– Robustness Checks

	Any moderate-heavy helminth infection, 1999				Moderate-heavy schistomiasis infection, 1999			
	Probit	OLS, spatial s.e.	Probit	Probit (Group 1 only)	Probit	OLS, spatial s.e.	Probit	Probit (Group 1 only)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Indicator for Group 1 (1998 Treatment) School	-0.31*** (0.06)	-0.28*** (0.06)	-0.32*** (0.06)		-0.09** (0.04)	-0.12* (0.06)	-0.07 (0.04)	
Group 1 pupils within 3 km (per 1000 pupils)	-0.21** (0.10)	-0.20** (0.09)		-0.28*** (0.08)	-0.12*** (0.05)	-0.17*** (0.04)		-0.06* (0.03)
Group 1 pupils within 3-6 km (per 1000 pupils)	-0.05 (0.08)	-0.11 (0.07)		-0.02 (0.06)	-0.15*** (0.04)	-0.14* (0.07)		-0.06*** (0.02)
Total pupils within 3 km (per 1000 pupils)	0.05 (0.04)	0.05 (0.06)	0.00 (0.04)	0.02 (0.02)	0.08*** (0.02)	0.12*** (0.04)	0.06*** (0.02)	0.01 (0.01)
Total pupils within 3-6 km (per 1000 pupils)	-0.02 (0.04)	0.02 (0.05)	-0.05* (0.03)	-0.02 (0.02)	0.04* (0.02)	0.04 (0.04)	-0.01 (0.02)	0.01 (0.01)
(Group 1 pupils within 3 km) / (Total pupils within 3 km)			-0.21* (0.12)				-0.10 (0.09)	
(Group 1 pupils within 3-6 km) / (Total pupils within 3-6 km)			-0.10 (0.23)				-0.46*** (0.12)	
Any moderate-heavy helminth infection, 1998				0.25*** (0.03)				
Moderate-heavy schistosomiasis infection, 1998								0.23** (0.10)
Grade indicators, school assistance controls, district exam score control	Yes	No	Yes	Yes	Yes	No	Yes	Yes
R^2	-	0.46	-	-	-	0.48	-	-
Root MSE	-	0.200	-	-	-	0.169	-	-
Number of observations	2330 (pupils)	49 (schools)	2330 (pupils)	603 (pupils)	2330 (pupils)	49 (schools)	2330 (pupils)	604 (pupils)
Mean of dep variable	0.41	0.41	0.41	0.25	0.16	0.16	0.16	0.08

Note: Grade 3-8 pupils. Robust standard errors in parentheses. Disturbance terms are clustered within schools for regressions 1, 3, 4, 5 and 7. Disturbance terms are allowed to be correlated across spaces using the method in Conley (1999) in regressions 2 and 6. Observations are weighted by total school population. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. The 1999 parasitological survey data are for Group 1 and Group 2 schools. The pupil population data is from the 1998 School Questionnaire. We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools.

Table A4: IV estimates of health and school participation externalities

	Any moderate-heavy helminth infection, January - March 99		Average individual school participation, May 98-March 99	
	Probit (1)	IV-2SLS (2)	OLS (3)	IV-2SLS (4)
Indicator for Group 1 (1998 Treatment) School	-0.18** (0.07)	-0.07 (0.10)	0.056*** (0.020)	0.024 (0.027)
Group 1 pupils within 3 km (per 1000 pupils)	-0.22** (0.11)	-0.19** (0.09)	0.022 (0.032)	0.019 (0.032)
Group 1 pupils within 3-6 km (per 1000 pupils)	-0.04 (0.08)	-0.03 (0.07)	-0.067*** (0.020)	-0.065*** (0.020)
Total pupils within 3 km (per 1000 pupils)	0.05 (0.04)	0.05 (0.03)	-0.040** 0.016)	-0.037** (0.017)
Total pupils within 3-6 km (per 1000 pupils)	-0.03 (0.04)	-0.02 (0.04)	0.035*** (0.011)	0.034 (0.011)
Indicator received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)	-0.06* (0.03)	-0.06 (0.05)	0.104*** (0.014)	0.022 (0.031)
(First year as treatment school Indicator)* (Received treatment, when offered)	-0.15** (0.06)	-0.26** (0.12)	-0.016 (0.020)	0.056 (0.045)
Grade indicators, school assistance controls, district exam score control	Yes	Yes	Yes	Yes
Time controls	No	No	Yes	Yes
R^2	-	-	0.37	-
Root MSE	-	0.450	0.217	0.218
Number of observations	2329	2329	18215	18215
Mean of dependent variable	0.41	0.41	0.793	0.793

Note: Disturbance terms are clustered within schools. Robust standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. The two instrumental variables are an indicator for girls under age 13 and all boys (ELG), and (ELG)*(Group 1 indicator). The coefficient on the Group 1 school indicator variable serves as an estimate of the within-school externality effect in 1998. This IV approach could overestimate the treatment effect if the treatment effect is heterogeneous, with sicker pupils benefiting most from treatment, and if among the girls over 13, the sickest girls are most likely to be treated in treatment schools. However, among the sub-sample of older girls, the compliance rate was not significantly related to infection status in 1998 (Table 6), and in 1999 under ten percent of older girls were treated (Table 3). We find similar effects even when we exclude the schools near the lake where older girls were likely to be treated (results not shown). Note that the IV estimates of within-school participation externalities should be interpreted as local average treatment effects for the older girls. Since school participation treatment effects are largest for younger pupils, it is not surprising that the IV externality estimates among the older girls are smaller than the OLS estimates, which are for the entire population. We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools.

Appendix B: Updated and preferred Miguel and Kremer (2004) tables

This appendix includes the relevant tables from Miguel and Kremer (2004), updated to use the final versions of all datasets, which contain our "preferred" analysis. As we argue in the main text of this note, it is not possible to precisely estimate externalities out to 6 km in this study. Thus, this set of tables includes externalities only out to a distance of 3 km. This change affects Tables I, VII, IX, and X.

Table I: 1998 Average pupil and school characteristics, pre-treatment

	Group 1 (25 schools)	Group 2 (25 schools)	Group 3 (25 schools)	G1–G3	G2–G3
Panel A: Pre-school to Grade 8					
Male	0.53	0.51	0.52	0.01	-0.01
				(0.02)	(0.02)
Proportion girls < 13 years, and all boys	0.89	0.89	0.88	0.00	0.01
				(0.01)	(0.01)
Grade progression (= Grade – (Age – 6))	-2.0	-1.8	-2.0	-0.0	0.1
				(0.1)	(0.1)
Year of birth	1986.2	1986.5	1985.8	0.4**	0.8***
				(0.2)	(0.2)
Panel B: Grades 3 to 8					
Attendance recorded in school registers (during 4 weeks prior to pupil survey)	0.973	0.963	0.969	0.003	-0.006
				(0.004)	(0.004)
Access to latrine at home	0.82	0.81	0.82	0.00	-0.01
				(0.03)	(0.03)
Have livestock (cows, goats, pigs, sheep) at home	0.66	0.67	0.66	-0.00	0.01
				(0.03)	(0.03)
Weight-for-age Z-score (low scores denote undernutrition)	-1.39	-1.40	-1.44	0.05	0.04
				(0.05)	(0.05)
Blood in stool (self-reported)	0.26	0.22	0.19	0.07**	0.03
				(0.03)	(0.03)
Sick often (self-reported)	0.10	0.10	0.08	0.02	0.02*
				(0.01)	(0.01)
Malaria/fever in past week (self-reported)	0.37	0.38	0.40	-0.03	-0.02
				(0.03)	(0.03)
Clean (observed by field workers)	0.60	0.66	0.67	-0.07**	-0.01
				(0.03)	(0.03)
Panel C: School characteristics					
District exam score 1996, grades 5-8‡	-0.10	0.09	0.01	-0.11	0.08
				(0.12)	(0.12)
Distance to Lake Victoria	10.0	9.9	9.5	0.6	0.5
				(1.9)	(1.9)
Pupil population	392.7	403.8	375.9	16.8	27.9
				(57.6)	(57.6)
School latrines per pupil	0.007	0.006	0.007	0.001	-0.000
				(0.001)	(0.001)
Proportion moderate-heavy infections in zone	0.37	0.37	0.36	0.01	0.01
				(0.03)	(0.03)
Group 1 pupils within 3 km††	430.4	433.2	344.5	85.9	88.7
				(116.2)	(116.2)
Total primary school pupils within 3 km	1272.7	1369.1	1151.9	120.8	217.2
				(208.1)	(208.1)

Note: School averages weighted by pupil population. Standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. Data from the 1998 ICS Pupil Namelist, 1998 Pupil Questionnaire and 1998 School Questionnaire.

‡1996 District exam scores have been normalized to be in units of individual level standard deviations, and so are comparable in units to the 1998 and 1999 ICS test scores (under the assumption that the decomposition of test score variance within and between schools was the same in 1996, 1998, and 1999).

†† This includes girls less than 13 years old, and all boys (those eligible for deworming in treatment schools).

Table VII: Deworming health externalities within and across schools, January to March 1999

	Any moderate-heavy helminth infection, 1999			Moderate-heavy schistosomiasis infection, 1999			Moderate-heavy geohelminth infection, 1999		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Indicator for Group 1 (1998 Treatment) School	-0.33***	-0.20***	-0.24***	-0.12***	-0.08	-0.10*	-0.29***	-0.18***	-0.22***
	(0.05)	(0.07)	(0.06)	(0.04)	(0.05)	(0.06)	(0.04)	(0.06)	(0.05)
Group 1 pupils within 3 km (per 1000 pupils)	-0.23**	-0.25**	-0.14	-0.13**	-0.13**	-0.10	-0.14	-0.15	-0.07
	(0.10)	(0.10)	(0.12)	(0.05)	(0.05)	(0.08)	(0.09)	(0.10)	(0.12)
Total pupils within 3 km (per 1000 pupils)	0.07*	0.08**	0.07**	0.10***	0.10***	0.10***	-0.01	-0.00	-0.01
	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)		-0.06**			0.04*			-0.10***	
		(0.03)			(0.02)			(0.03)	
(Group 1 Indicator) * Received treatment, when offered		-0.14**			-0.05			-0.11**	
		(0.07)			(0.04)			(0.05)	
(Group 1 Indicator) * Group 1 pupils within 3 km (per 1000 pupils)			-0.23*			-0.06			-0.18
			(0.13)			(0.08)			(0.12)
Grade indicators, school assistance controls, district exam score control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	2330	2329	2330	2330	2329	2330	2330	2329	2330
Mean of dependent variable	0.41	0.41	0.41	0.16	0.16	0.16	0.32	0.32	0.32

Note: Grade 3-8 pupils. Probit estimation, robust standard errors in parentheses. Disturbance terms are clustered within schools. Observations are weighted by total school population. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. The 1999 parasitological survey data are for Group 1 and Group 2 schools. The pupil population data is from the 1998 School Questionnaire. The geohelminths are hookworm, roundworm, and whipworm. We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools.

Table IX: School participation, direct effects and externalities
Dependent variable: Average individual school participation, by year

	OLS (1)	OLS (2)	OLS (3)	OLS (4) May 98-March 99	OLS (5) May 98-March 99	OLS (6) May 98-March 99	IV-2SLS (7) May 98-March 99
Moderate-heavy infection, early 1999						-0.028*** (0.009)	-0.282** (0.111)
Treatment school (T)	0.057*** (0.014)						
First year as treatment school (T1)		0.063*** (0.015)	0.065*** (0.014)	0.062*** (0.022)	0.044* (0.024)		
Second year as treatment school (T2)		0.039* (0.021)	0.036* (0.021)				
Treatment school pupils within 3 km (per 1000 pupils)			0.046** (0.022)		0.027 (0.040)		
Total pupils within 3 km (per 1000 pupils)			-0.031** (0.013)		-0.034* (0.019)	0.016 (0.015)	-0.032* (0.017)
Indicator received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)					0.104*** (0.014)		
(First year as treatment school Indicator)* (Received treatment, when offered)					-0.013 (0.020)		
1996 district exam score, school average	0.071*** (0.021)	0.070*** (0.021)	0.070*** (0.022)	0.058* (0.032)	0.060* (0.031)	0.016 (0.024)	-0.004 (0.021)
Grade indicators, school assistance controls, and time controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.22	0.22	0.22	0.33	0.36	0.28	-
Root MSE	0.279	0.279	0.278	0.223	0.218	0.150	0.071
Number of observations	56496	56496	56496	18215	18215	2327	49 (schools)
Mean of dependent variable	0.747	0.747	0.747	0.793	0.793	0.884	0.884

Note: The dependent variable is average individual school participation in each year of the program (Year 1 is to March 1999, and Year 2 is May 1999 to November 1999); disturbance terms are clustered within schools. Robust standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. Additional explanatory variables include an indicator variable for girls < 13 years and all boys, and the rate of moderate-heavy infections in geographic zone, by grade (zonal infection rates among grade 3 and 4 pupils are used for pupils in grades 4 and below and for pupils initially recorded as drop-outs as there is no parasitological data for pupils below grade 3; zonal infection rates among grade 5 and 6 pupils are used for pupils in grades 5 and 6, and similarly for grades 7 and 8). Participation is computed among all pupils enrolled at the start of the 1998 school year. Pupils present during an unannounced NGO school visit are considered participants. Pupils had approximately 3.8 attendance observations per year. Regressions 6 and 7 include pupils with parasitological information from early 1999, restricting the sample to a random subset of Group 1 and Group 2 pupils. The number of treatment school pupils from May 1998 to March 1999 is the number of Group 1 pupils, and the number of treatment school pupils after March 1999 is the number of Group 1 and Group 2 pupils. The instrumental variables in regression 7 are the Group 1 (treatment) indicator variable, Treatment school pupils within 3 km, and the remaining explanatory variables. We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools.

Table X: Academic examinations, individual-level data

	Dependent variable: ICS Exam Score (normalized by standard)		
	(1)	(2)	(3) Among those who filled in the 1998 pupil survey
Average school participation (during the year of the exam)	0.63***		
	(0.07)		
First year as treatment school (T1)		-0.042	-0.043
		(0.048)	(0.051)
Second year as treatment school (T2)		-0.014	-0.011
		(0.075)	(0.085)
1996 District exam score, school average	0.74***	0.75***	0.78***
	(0.07)	(0.06)	(0.07)
Grade indicators, school assistance controls, and local pupil density controls	Yes	Yes	Yes
R^2	0.14	0.13	0.14
Root MSE	0.919	0.924	0.918
Number of observations	24979	24979	19072
Mean of dependent variable	0.019	0.019	0.039

Note: Each data point is the individual-level exam result in a given year of the program (either 1998, or 1999); disturbance terms are clustered within schools. Linear regression, robust standard errors in parentheses. Significantly different than zero at 99 (***), 95 (**), and 90 (*) percent confidence. Regression 3 includes only pupils who completed the 1998 Pupil Questionnaire. Additional explanatory variables include an indicator variable for girls < 13 years and all boys, and the rate of moderate-to-heavy infections in geographic zone, by grade (zonal infection rates among grade 3 and 4 pupils are used for pupils in grades 4 and below and for pupils initially recorded as dropouts as there is no parasitological data for pupils below grade 3; zonal infection rates among grade 5 and 6 pupils are used for pupils in grades 5 and 6, and similarly for grades 7 and 8). The local pupil density terms include treatment school pupils within 3 km (per 1000 pupils), and total pupils within 3 km (per 1000 pupils). We use the number of girls less than 13 years old and all boys (the pupils eligible for deworming in the treatment schools) as the school population for all schools. The ICS tests for 1998 and 1999 were similar in content, but differed in two important respects. First, the 1998 exam featured multiple-choice questions while the 1999 test featured short answers. Second, while each grade in 1998 was administered a different exam, in 1999 the same exam – featuring questions across a range of difficulty levels – was administered to all pupils in grades 3 to 8. Government district exams in English, Maths, Science-Agriculture, Kiswahili, Geography-History, Home Science, and Arts-Crafts were also administered in both years. Treatment effect estimates are similar for both sets of exams (results not shown).