
Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities

Guide to Replication of Miguel and Kremer (2004)

Data released to public: January 2007

Edward Miguel
Department of Economics
University of California, Berkeley
(emiguel@econ.berkeley.edu)

Michael Kremer
Department of Economics
Harvard University
(mkremer@fas.harvard.edu)

Table of Contents

1. Introduction.....	3
2. Replication Files	3
3. Note on Data Sets Used in Analysis.....	3
4. Typographical and Other Errors in the Published Paper	4
4.1 Rounding off errors	4
4.2 Statistical significance errors (wrong number of stars)	6
4.3 STATA coding errors	7
4.3.1 Moderate-heavy geohelminth infection, 1999	7
4.3.2 Local school and pupil population densities	7
4.3.3 Table 9, “Treatment School Pupils within 3-6 km”	8
4.4 Other typos or errors.....	8
5. Appendix	11

1. Introduction

In order to allow for replication of the analysis provided in Miguel and Kremer (2004), the data and do-files employed in the published paper have been made publicly available. This document provides notes on the replication itself. Background to the study, as well as notes on the data sets, can be found in companion documents entitled “Data User’s Guide” and “Codebooks”.

2. Replication Files

Each table in Miguel and Kremer (2004) can be replicated using a single corresponding do file. Additional “updated” do-files exist for tables 7, 8, and 9. This is explained in more detail below.

3. Note on Data Sets Used in Analysis

In recently revisiting the do files to prepare this document, we discovered that several results in the published version of the paper were produced using “intermediate”, or earlier, versions of the data sets rather than the final version. 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 two data sets, versions that were progressively cleaner over time. Cleaning typically took the form of eliminating duplicate observations that had found their way into the data, correcting data entry errors through hard copy checks, and better matching across files. While the bulk of results in the paper use the “final” versions of the data, not all tables were completely updated during the journal revision process and some tables instead use “intermediate” versions of the data. The important thing is that the extent of data cleaning was only moderate, so that using versions of the data from 2000, 2001, and 2002, say, leads to almost identical results.

Specifically, the published versions of Table 7 and Appendix Tables A2 and A3 use an intermediate version of the “comply.dta” data set, Table 8 uses an intermediate version of the “namelist.dta” data set, and Table 9 and Appendix Table A4 use intermediate versions of both the “comply.dta” and the “namelist.dta” data sets. Two sets of do files are provided for each of these tables – one which replicates the original published results (using some “intermediate”, or not fully cleaned, data sets), and a second which updates these results using the “final” version of the data sets.

An updated version of the tables in Miguel and Kremer (2004) is presented in the appendix to this document. These tables were produced using the “final” versions of all data sets, and corrected the errors noted in section 4 below. These updated tables highlight that there is little substantive difference between those results presented in the published paper and the updated versions presented here, since the intermediate and final data sets are so similar.

4. Typographical and Other Errors in the Published Paper

Several typographical errors have also been discovered in the tables presented in Miguel and Kremer (2004). In what follows, we detail these errors in four categories: (i) rounding off errors, (ii) statistical significance errors (wrong number of stars), (iii) STATA coding errors, and (iv) other typos or errors. Rounding off errors account for the bulk of the typographical errors in the paper, and we start with these.

Once again, these errors affect little of the published paper's substantive results. An updated version of the tables in Miguel and Kremer (2004), incorporating the changes noted in sections 3 and 4, is presented in the appendix to this document.

4.1 Rounding off errors

Rounding off errors come in two varieties. First, we discovered that a fair number of estimates were "truncated" rather than rounded off (e.g., 0.787 became 0.78 rather than 0.79). Second, we found that several estimates were incorrectly rounded off when estimates were shortened, usually from three decimal points to two decimal points. In earlier versions of the paper (including the NBER Working Paper version, Miguel and Kremer 2001), a number of tables presented statistics with three decimal points, while the published version often uses two decimal points. In the process of shortening these figures, a number of errors were made. Consider the first entry of Table 3 (the upper left corner entry), 0.78. It turns out that the true value of this statistic is 0.7745. In an earlier version which presented data with three decimal points, this was presented as 0.775. When we shifted to two decimal points, this figure was incorrectly rounded up to 0.78 when it should have been rounded down to 0.77.

In all these cases, the rounding errors obviously tend to lead to very minor changes in estimates. These errors are detailed below.

Table 2:

At least one infection, born since 1985, prevalence of infection: 0.93 (published version 0.92).

Table 3:

TOP PANEL:

Any medical treatment in 1998, Group 1, eligibles: 0.77 (published version 0.78)

Any medical treatment in 1998, Group 1, ineligibles: 0.20 (published version 0.19)

Round 1 Albendazole 1998, Group 1, eligibles: 0.68 (published version 0.69)

MIDDLE PANEL:

Any medical treatment in 1999, Group 1, eligibles: 0.58 (published version 0.59)

Any medical treatment in 1999, Group 2, eligibles: 0.54 (published version 0.55)

Any medical treatment in 1999, Group 2, ineligibles: 0.09 (published version 0.10)

Round 1 Albendazole 1999, Group 2, ineligibles: 0.05 (published version 0.06)

Round 1 Praziquantel 1999, Group 3, eligibles: 0.00 (published version 0.01)

Round 2 Albendazole 1999, Group 1, eligibles: 0.52 (published version 0.53)

Round 2 Albendazole 1999, Group 2, eligibles: 0.50 (published version 0.51)

Round 2 Albendazole 1999, Group 2, ineligibles: 0.07 (published version 0.08)

BOTTOM PANEL:

Any medical treatment in 1999, Group 2, ineligibles: 0.14 (published version 0.13)
Round 1 Praziquantel 1999, Group 1, eligibles: 0.54 (published version 0.53)
Round 1 Praziquantel 1999, Group 1, ineligibles: 0.08 (published version 0.07)
Round 1 Praziquantel 1999, Group 2, eligibles: 0.46 (published version 0.45)
Round 1 Praziquantel 1999, Group 3, eligibles: 0.00 (published version 0.01)

Table 4:

Total transfers in Group 1 schools, 1998: 0.020 (published version 0.021)
Total transfers in Group 3 schools, 1998: 0.020 (published version 0.021)

Table 5:

PANEL B:

Sick in past week, 1999, Group 1: 0.40 (published version 0.41)
Sick in past week, 1999, coefficient on Group 1 - Group 2: -0.05 (published version -0.04)
Height-for-age Z-score, 1999, coefficient on Group 1 - Group 2: 0.08 (published version 0.09)
Hemoglobin concentration, 1999, Group 1: 124.9 (published version 124.8)
Hemoglobin concentration, 1999, Group 2: 123.3 (published version 123.2)

Table 6:

PANEL A:

Proportion of 1998 parasitological sample tracked in 1999 sample, Group 1, Untreated: 0.35
(published version 0.36)
Access to latrine at home, 1998 for Group 1, Treated: 0.85 (published version 0.84)

PANEL B:

Hookworm moderate-heavy infection, 1999, coefficient on Group 1 untreated 1998 – Group 2
untreated 1999: -0.10 (published version -0.09)
Any moderate-heavy infection, 1998 for Group 1, Untreated Girls ≥ 13 years: 0.30 (published
version 0.28)
Any moderate-heavy infection, 1999 for Group 1, Untreated Girls ≥ 13 years: 0.44 (published
version 0.43)
Any moderate-heavy infection, 1999, coefficient on Group 1 Untreated 1998 – Group 2
Untreated 1999: -0.09 (published version -0.10)

Table 7:

Column (2), coefficient on “indicator for Group 1 (1998 treatment) school”: -0.13 (published
version -0.12)
Column (5), standard error on “received first year of deworming treatment, when offered”: 0.01
(published version 0.02)

Table 9:

Column (4), coefficient on “first year as treatment school”: 0.063 (published version 0.062)
Column (4), root MSE: 0.224 (published version 0.223)
Column (5), coefficient on “1996 district exam score, school average”: 0.092 (published version
0.091)

Column (7), coefficient on “moderate-heavy infection, early 1999”: -0.204 (published version -0.203)

Column (7), coefficient on “1996 district exam score”: 0.004 (published version 0.003)

Column (7), standard error on “1996 district exam score”: 0.022 (published version 0.023)

Table 10:

Column (1), mean of dependent variable: 0.019 (published version 0.020)

Column (2), standard error on “second year as treatment school”: 0.072 (published version 0.071)

Column (2), mean of dependent variable: 0.019 (published version 0.020)

Column (3), coefficient on “second year as treatment school”: 0.008 (published version 0.009)

Appendix Table A4:

Column (1), coefficient on “indicator for group 1 school”: -0.13 (published version -0.12)

Column (2), coefficient on “indicator received first year of deworming treatment, when offered”: -0.05 (published version -0.06)

Column (4), standard error on “total pupils within 3-6 km”: 0.026 (published version 0.027)

4.2 Statistical significance errors (wrong number of stars)

In several cases the published paper either failed to note statistical significance or overstated it.

Table 1:

PANEL B:

Sick often (self-reported), Group 1 - Group 3 difference is not statistically significant

Sick often (self-reported), Group 2 - Group 3 difference is statistically significant at 90% confidence

Table 5:

Proportion anemic, 1999, Group 1 - Group 2 difference is not statistically significant

Table 6:

PANEL A:

Grade progression, 1998, Group 1 Treated 1998 - Group 2 Treated 1998 difference is not statistically significant.

PANEL C:

School participation rate, May 1998 to March 1999, Group 1 Treated 1998 – Group 2 Treated 1999 difference is statistically significant at 90% confidence

Table 7:

Column (2), interaction between Group 1 indicator and received treatment when offered is statistically significant at 95% confidence

Table 8:

PANEL A:

Girls \geq 13 years, Group 1 – Groups 2&3 difference is statistically significant at 90% confidence

Table 9:

Column (1), "treatment school" is statistically significant at 95% confidence
Column (3), "second year as treatment school" is not statistically significant
Column (4), "first year as treatment school" is statistically significant at 99% confidence.
Column (4), "1996 district exam score, school average" is significant at 90% confidence.
Column (7), "moderate-heavy infection, early 1999" is significant at 95% confidence

Appendix Table A4:

Column (1), "first year as treatment school indicator * received treatment, when offered" is statistically significant at 90% confidence

4.3 STATA coding errors**4.3.1 Moderate-heavy geohelminth infection, 1999**

Construction of the original "moderate-heavy geohelminth infection, 1999" variable contained a slight coding error. This mistake affects results in Table 7, although it does not change the substantive findings. The variable "any_geo99_original" is the original version containing the coding error, while the variable "any_geo99_updated" contains the corrected data. Results both with the coding mistake and with the correction are contained in the STATA output log files for Table 7 (table7-log_jan07.txt) where noted. The updated results are also shown in the appendix to this document.

4.3.2 Local school and pupil population densities

Local school and pupil population figures are important in the estimation of cross-school externalities in Miguel and Kremer (2004). The data user will notice that in the "schoolvar.dta" data set, for each measure of the number of children or schools located within a certain distance (measures typically named "sch#_#km" or "pop#_#km" in the schoolvar.dta data set), there exist two variables. One variable has the suffix "_original" while the other variable has the suffix "_updated". The variables denoted "_original" were those used in the analysis in Miguel and Kremer (2004). However, during the preparation of this replication documentation, two small coding errors were discovered in the creation of these original local density figures.

One coding error truncated the number of schools that were counted in the school and population densities to twelve, rather than allowing all 74 other schools to be included in this count. Since there were fewer than 12 schools located at distances of up to four kilometers from any given PSDP school, this coding error does not affect school and population density figures in the published paper for distances of 1-3 kilometers. However, density figures for distances of 3-6 kilometers do change somewhat.

A second coding error was discovered that miscalculated local density figures for just three schools – identification numbers 108, 109, and 115. This coding error essentially miscalculated densities in treatment group 1 for school 108, treatment group 2 for school 109, and treatment group 3 for school 115.

Due to these errors, use of these “_original” variables in future analysis is not recommended. The variables with the “_updated” suffix eliminate this coding error, and also employ a Geographic Information System (GIS) computer package to convert the GPS information into more precise measures of distance. The original variables were generated manually using a simple mathematical formula to convert GPS figures into distances. The updated variables now provide a more precise measure of local school and pupil population densities.

Results in the paper do change slightly once these coding errors in the local pupil density figures are fixed, and the data is updated to incorporate more sophisticated conversions between GPS and distance estimates. The tables in the appendix to this document show the updated results.

4.3.3 Table 9, “Treatment School Pupils within 3-6 km”

A coding error was discovered in the calculation of a variable used only in column (3) of Table 9, “treatment school pupils within 3-6 km.” This error summed Group 1 pupils within 0-3 kilometers and Group 2 pupils within 3-6 kilometers, rather than summing Group 1 and Group 2 pupils both within 3-6 kilometers. The error was corrected in the code written to produce Table 9, and the only substantive change in results was a loss of the marginal significance on the “second year as treatment school (T2)” term.

4.4 Other typos or errors

The remaining errors appear to be simple typos, or are due to our failure to fully update all empirical results during the journal revision process, combined with the use of “intermediate” versions of the data in earlier paper drafts.

Table 2:

The figure listed as “at least two infections, prevalence of infection” in fact corresponds to “exactly two infections, prevalence of infection”. The figure for “at least two infections, prevalence of infection” would be 0.65.

The figure listed as “at least three infections, prevalence of infection” in fact corresponds to “exactly three infections, prevalence of infection”. The figure for “at least three infections, prevalence of infection” would be 0.34.

Table 5:

Obs. for parasitological results: 1,466 for Group 2 (published version 1,467)

Obs. for hemoglobin results: 769 total (published version 778), 290 for Group 1 (published version 292), 479 for Group 2 (published version 486)

Obs. for 1999 Pupil Questionnaire health outcomes: 9,039 total (published version 9,102), 3,545 for Group 1 (published version 3,562), 5,497 for Group 2 and Group 3 (published version 5,540)

Table 6:

The difference between “malaria/fever in past week (self-reported), 1998” for Group 1 untreated 1998 and Group 2 untreated 1999 has a coefficient of 0.02 (published version -0.01)

The difference between “whipworm moderate-heavy infection, 1999” for Group 1 Treated 1998 and Group 2 Treated 1999 has standard error 0.05 (published version 0.16)

The “school participation rate, May 1998 to March 1999” for group 1 untreated in 1998 is 0.774 (published version 0.764)

The “school participation rate, May 1998 to March 1999” for group 2 untreated in 1998 is 0.690 (published version 0.684)

The difference between “school participation rate, May 1998 to March 1999” for group 1 treated 1998 and group 2 treated 1999 has a standard error of 0.033 (published version 0.032)

The difference between “school participation rate, May 1998 to March 1999” for group 1 untreated 1998 and group 2 untreated 1999 has a coefficient of 0.084 (published version 0.080) and a standard error of 0.037 (published version 0.039)

Obs. for the parasitological survey: 669 Group 1 treated 1998 (published version 670), 76 Group 1 untreated 1998 (published version 77), 874 Group 2 treated 1999 (published version 873), 349 Group 2 untreated 1999 (published version 352)

Table 8:

PANEL A

Row 3: Column (1) 0.797 (published version 0.795), Column (2) 0.689 (published version 0.688), Column (3) 0.707 (published version 0.703)

PANEL B

Row 2: Column (1) 0.670 (published version 0.627), Column (2) 0.678 (published version 0.649), Column (3) 0.619 (published version 0.588), Column (4) 0.050* (s.e. 0.030) (published version 0.039 (s.e. 0.035)), Column (5) 0.058* (s.e. 0.030) (published version 0.061* (s.e. 0.035))

Table 9:

Column (4), Mean of dependent variable: 0.793 (published version 0.784)

Column (5), Mean of dependent variable: 0.793 (published version 0.784)

Column (7), Root MSE: 0.069 (published version 0.773)

Table 10:

Columns (1) and (2), Number of observations: 24979 (published version 24958)

Appendix Table A2:

Column (1), standard error of “treatment school pupils within 3 km” is 0.08 (published version 0.06)

Column (1), standard error of “treatment school pupils within 3-6 km” is 0.06 (published version 0.07)

Column (1), standard error of “total pupils within 3 km” is 0.06 (published version 0.05)

Column (1), standard error of “total pupils within 3-6 km” is 0.04 (published version 0.06)

Column (1), mean of dependent variable is 0.76 (published version 0.66)

Column (2), standard error of “treatment school pupils within 3-6 km” is 0.06 (published version 0.05)

Column (2), standard error of “total pupils within 3 km” is 0.07 (published version 0.08)

Appendix Table A3:

Column (8), standard error on “group 1 pupils within 3 km”: 0.03 (published version 0.02)
Column (8), “Group 1 pupils within 3 km” is statistically significant at 95% confidence
Column (8), coefficient on “group 1 pupils within 3-6 km”: -0.06 (published version -0.05)
Column (8), standard error on “group 1 pupils within 3-6 km”: 0.02 (published version 0.01)
Column (8), standard error on “total pupils within 3 km”: 0.02 (published version 0.01)
Column (8), “total pupils within 3 km” is statistically significant at 90% confidence
Column (8), coefficient on “total pupils within 3-6 km”: 0.03 (published version 0.04)
Column (8), standard error on “total pupils within 3-6 km”: 0.02 (published version 0.01)
Column (8), “total pupils within 3-6 km” is statistically significant at 95% confidence
Column (8), “moderated-heavy schistosomiasis infection, 1998” is statistically significant at 95% confidence
Columns (1), (3), (5) and (7), number of observations: 2328 (published version 2326)

Appendix Table A4:

Columns (1) and (2), number of observations: 2328 (published version 2326)
Columns (3) and (4), mean of dependent variable: 0.793 (published version 0.784)

5. Appendix

This appendix includes all tables in Miguel and Kremer (2004), updated to use the “final” (updated) versions of all datasets and corrected of all rounding, significance level, typographical and coding errors. These tables show little substantive differences between the published and updated results. More specifically, the results in tables I through VI, VIII, X and A2 are almost identical to the published paper. Tables VII, IX, A3 and A4, using the updated data and local density calculations as described above, continue to show the strong deworming treatment effect on eradicating helminth infections and improving school participation, as well as the positive externalities of other treated pupils within three kilometers, although the positive externalities of pupils at greater distances are somewhat weaker.

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

	Group 1 (25 schools)	Group 2 (25 schools)	Group 3 (25 schools)	Group 1 – Group 3	Group 2 – Group 3
<i>Panel A: Pre-school to Grade 8</i>					
Male	0.53	0.51	0.52	0.01 (0.02)	-0.01 (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))	-2.1	-1.9	-2.1	-0.0 (0.1)	0.1 (0.1)
Year of birth	1986.2	1986.5	1985.8	0.4 ^{**} (0.2)	0.8 ^{***} (0.2)
<i>Panel B: Grades 3 to 8</i>					
Attendance recorded in school registers (during the four weeks prior to the pupil survey)	0.973	0.963	0.969	0.003 (0.004)	-0.006 (0.004)
Access to latrine at home	0.82	0.81	0.82	0.00 (0.03)	-0.01 (0.03)
Have livestock (cows, goats, pigs, sheep) at home	0.66	0.67	0.66	-0.00 (0.03)	0.01 (0.03)
Weight-for-age Z-score (low scores denote undernutrition)	-1.39	-1.40	-1.44	0.05 (0.05)	0.04 (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 (0.01)	0.02 [*] (0.01)
Malaria/fever in past week (self-reported)	0.37	0.38	0.40	-0.03 (0.03)	-0.02 (0.03)
Clean (observed by field workers)	0.60	0.66	0.67	-0.07 ^{**} (0.03)	-0.01 (0.03)
<i>Panel C: School characteristics</i>					
District exam score 1996, grades 5-8 [‡]	-0.10	0.09	0.01	-0.11 (0.12)	0.08 (0.12)
Distance to Lake Victoria	10.0	9.9	9.5	0.6 (1.9)	0.5 (1.9)
Pupil population	392.7	403.8	375.9	16.8 (57.6)	27.9 (57.6)
School latrines per pupil	0.007	0.006	0.007	0.001 (0.001)	-0.000 (0.001)
Proportion moderate-heavy infections in zone	0.37	0.37	0.36	0.01 (0.03)	0.01 (0.03)
Group 1 pupils within 3 km ^{††}	430.4	433.2	344.5	85.9 (116.2)	88.7 (116.2)
Group 1 pupils within 3-6 km	1157.6	1043.0	1297.3	-139.7 (199.3)	-254.4 (199.3)
Total primary school pupils within 3 km	1272.7	1369.1	1151.9	120.8 (208.1)	217.2 (208.1)
Total primary school pupils within 3-6 km	3431.3	3259.8	3502.1	-70.8 (366.0)	-242.3 (366.0)

[†]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 infection intensity, 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	-

[†]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 years, and all boys	Girls ≥ 13 years	Girls < 13 years, and all boys	Girls ≥ 13 years	Girls < 13 years, and all boys	Girls ≥ 13 years
	<i>Treatment</i>		<i>Comparison</i>		<i>Comparison</i>	
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
	<i>Treatment</i>		<i>Treatment</i>		<i>Comparison</i>	
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

[†]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 treatment threshold (30 percent prevalence) in that year.

Table IV: Proportion of pupil transfers across schools

School in early 1998 (pre-treatment)	1998 transfer to a			1999 transfer to a		
	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	Group 1 – Group 2
<i>Panel A: Helminth Infection Rates</i>			
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)
<i>Panel B: Other Nutritional and Health Outcomes</i>			
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)
<i>Panel C: Worm Prevention Behaviors</i>			
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)

[†]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, 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.

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

	Group 1, Treated in 1998	Group 1, Untreated in 1998	Group 2, Treated in 1999	Group 2, Untreated in 1999	(Group 1 Treated 1998) – (Group 2, Treated 1999)	(Group 1, Untreated 1998) – (Group 2, Untreated 1999)
<i>Panel A: Selection into Treatment</i>						
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)
<i>Panel B: Health Outcomes</i>						
<i>Girls < 13 years, and all boys</i>						
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)
<i>Girls ≥ 13 years</i>						
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)
<i>Panel C: School Participation</i>						
School participation rate, May 1998 to March 1999 ^{††}	0.872	0.774	0.808	0.690	0.064* (0.033)	0.084** (0.037)

[†]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*** (0.06)	-0.18** (0.07)	-0.21* (0.11)	-0.09** (0.04)	-0.06 (0.05)	-0.03 (0.06)	-0.30*** (0.05)	-0.19*** (0.06)	-0.26*** (0.09)
Group 1 pupils within 3 km (per 1000 pupils)	-0.21** (0.10)	-0.22** (0.11)	-0.10 (0.14)	-0.12** (0.05)	-0.12** (0.05)	-0.08 (0.07)	-0.12 (0.09)	-0.13 (0.10)	-0.06 (0.12)
Group 1 pupils within 3-6 km (per 1000 pupils)	-0.05 (0.08)	-0.04 (0.08)	-0.08 (0.11)	-0.15*** (0.04)	-0.15*** (0.04)	-0.13*** (0.05)	0.06 (0.06)	0.08 (0.06)	0.03 (0.09)
Total pupils within 3 km (per 1000 pupils)	0.05 (0.04)	0.05 (0.04)	0.05 (0.03)	0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)
Total pupils within 3-6 km (per 1000 pupils)	-0.02 (0.04)	-0.03 (0.04)	-0.02 (0.04)	0.04* (0.02)	0.04* (0.02)	0.04* (0.02)	-0.04 (0.03)	-0.05 (0.03)	-0.04 (0.03)
Received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)		-0.06* (0.03)			0.04** (0.02)			-0.10*** (0.03)	
(Group 1 Indicator) * Received treatment, when offered		-0.15** (0.06)			-0.04 (0.04)			-0.11** (0.05)	
(Group 1 Indicator) * Group 1 pupils within 3 km (per 1000 pupils)			-0.27* (0.14)			-0.07 (0.08)			-0.16 (0.11)
(Group 1 Indicator) * Group 1 pupils within 3-6 km (per 1000 pupils)			0.01 (0.09)			-0.03 (0.06)			0.03 (0.07)
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

[†]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)		
<i>Panel A: First year post-treatment (May 1998 to March 1999)</i>					
	<i>1st Year Treatment</i>	<i>Comparison</i>	<i>Comparison</i>	<i>Group 1 – (Groups 2 & 3)</i>	<i>Group 2 – Group 3</i>
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)
<i>Panel B: Second year post-treatment (March to November 1999)</i>					
	<i>2nd Year Treatment</i>	<i>1st Year Treatment</i>	<i>Comparison</i>	<i>Group 1 – Group 3</i>	<i>Group 2 – Group 3</i>
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)

[†]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

[†] 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 (0.047)	-0.036 (0.049)
Second year as treatment school (T2)		-0.015 (0.079)	-0.013 (0.088)
1996 District exam score, school average	0.74 ^{***} (0.07)	0.72 ^{***} (0.07)	0.75 ^{***} (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

[†] 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).

**Appendix 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 (0.07)	-0.04 (0.12)
Treatment school pupils within 3-6 km (per 1000 pupils)	0.08 (0.05)	-0.01 (0.06)
Total pupils within 3 km (per 1000 pupils)	0.09** (0.03)	0.04 (0.08)
Total pupils within 3-6 km (per 1000 pupils)	-0.03 (0.03)	0.00 (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

[†]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.

Appendix Table A3: Deworming health externalities– Robustness Checks[†]

	Any moderate-heavy helminth infection, 1999				Moderate-heavy schistosomiasis 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 dependent variable	0.41	0.41	0.41	0.25	0.16	0.16	0.16	0.08

[†]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.

Appendix 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

[†] 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.