

The long run effects of early childhood deworming on literacy and numeracy: Evidence from Uganda

Kevin Croke

Department of Global Health and Population, Harvard School of Public Health

kevinjcroke@gmail.com

This version July 17, 2014

Acknowledgements: Thanks to Owen Ozier, for suggesting the Uganda deworming project as a candidate for long run follow up and for providing the list of treatment and control parishes; to Harold Alderman and Günther Fink for valuable feedback; and to seminar participants at the Harvard School of Public Health.

Abstract: *This paper analyzes the long run impact of a cluster-randomized trial in eastern Uganda that provided mass deworming treatment to a sample of preschool aged children from 2000 to 2003. An early impact evaluation of this intervention found that the treatment group, comprised of children aged 1-7, showed increased weight gain compared to controls (Alderman et al. 2007). Since there is now a large literature linking early life health, often proxied by weight, to long run outcomes (including cognitive, educational, health, and labor market outcomes), I use data collected in these communities 7-8 years after the end of the deworming trial to see whether children in treatment communities have higher scores than children in control communities on simple numeracy and literacy tests. I find that children who lived in treatment communities during the period in question have test scores 0.2-0.4 standard deviations higher than those in control parishes. Effects are larger for math than for English literacy scores. The effect is robust to a wide range of alternate specifications and inclusion of socioeconomic control variables, and to a placebo treatment test. Girls and children from the poorest income quintiles experience relatively larger gains.*

1 Introduction

While mass provision of deworming treatment to school aged children living in highly endemic regions has long been a WHO-recommended policy, debate about the effectiveness of this intervention has recently emerged, tied to the 2012 publication of an updated Cochrane Systematic Review (Taylor-Robinson et al 2012) which finds essentially no evidence of benefit to mass deworming programs. This paper adds new evidence to the debate about the relationship between early childhood deworming and educational and cognitive outcomes. I analyze the long run impact of a cluster-randomized trial in eastern Uganda that provided mass deworming treatment to a sample of preschool aged children from 2000 to 2003. An early impact evaluation of this project found that the treatment group, comprised of children aged 1-7, showed increased weight gain compared to controls (Alderman et al. 2007). Since there is now a large literature linking early life health to long run outcomes (including cognitive, educational, health, and labor market outcomes), I use data collected in these communities 7-8 years after the end of the deworming trial to see whether children in treatment communities have higher scores than children in control communities on simple numeracy and literacy tests. I find that children who lived in treatment communities during the period in question have test scores 0.2-0.4 standard deviations higher than those in control parishes. Effects are larger for math than for English literacy scores. The effect is robust to a wide range of alternate specifications and inclusion of socioeconomic control variables, and to a placebo treatment test. When the clustering of errors is adjusted for the relatively small (22 total) number of clusters, the treatment effect on math results is unchanged, while the effect on literacy and combined scores are significant at the 10% level. As the following discussion will show, I cannot precisely identify the mechanism through which this gain is transmitted. However, I can rule out

any health differences that operate through a long run school enrollment or contemporaneous school attendance channels, since these do not differ between treatment and control communities.

2 Background on soil-transmitted helminths (intestinal worms)

As many as two billion people are estimated to suffer from intestinal worms (also known as soil-transmitted helminths) such as roundworm, hookworm, or whipworm (Bundy et al 2009). The largest burden of disease is found in sub-Saharan Africa and in South Asia. Worm infection, although rarely fatal, is associated with a wide range of health problems including stunted growth, reduced food absorption, loss of appetite, listlessness, and anemia. Among pre-school age children, worm infection is associated with slowed growth, and among school age children it has been linked to poor attendance at school and reduced performance on cognitive tasks. Soil-transmitted helminths are largely transmitted through fecal-oral contact, with the result that the disease burden is highest among populations with poor sanitation and inadequate access to health care, and in conditions of poverty. Inexpensive and effective treatment is available for worms, in the form of a single dose of drugs, such as albendazole or mebendazole. These drugs are safe, have minimal side effects, and are well tolerated by both infected and non-infected populations.

The availability of inexpensive treatment that has no significant side effects for uninfected children – and the fact that screening children for infection is significantly more expensive than treating them – has pointed global health policymakers towards a policy of mass treatment in high prevalence populations, such as school age children. In fact, the World Health Organization advocates for annual deworming of all children in regions where STH prevalence is over 20%, and twice annually where prevalence is above 50%. Several recent papers have found large benefits to interventions following this policy in Western Kenya. Miguel and Kremer (2004), for example, find large impacts of school-based mass deworming on school attendance in western Kenya, reducing absenteeism by 7.5 percentage points. Baird et al (2011) track these cohorts over time through the Kenya Life Panel Survey (KLPS), and find positive educational and labor market outcomes for the cohorts that were dewormed more frequently (the original study was a phase-in randomization so there is no pure control in the long run). Ozier (2012) returns to the original study schools approximately 10 years later and finds large spillover effects on children who were under one year of age when their community was dewormed, with increased scores on cognitive tests equivalent to a full year of schooling. Moving across the border from Western Kenya to the program in eastern Uganda that this paper focuses on, the short run impact evaluation mentioned above (Alderman et al 2007) found that beneficiaries of the deworming program (e.g those who attended at least two Child Health Days over the

program duration in the treatment group) gained 5-10% more in weight than Child Health Day attendees in the control communities, where deworming was not offered.¹

2a The current deworming controversy

The evidence from these studies, and from earlier biomedical trials, has led to various international expert groups to rank mass deworming very highly on a number of policy prioritization and cost effectiveness exercises.² For example, the 2012 Copenhagen Consensus ranked school-based deworming as its fourth most important priority for development funders.³ Similarly, researchers affiliated with MIT's Poverty Action Lab found that it was the second most cost effective way to increase school attendance, producing 13.9 additional years of schooling for every \$100 spent.⁴

These recommendations, however, are in tension with the findings from a series of Cochrane Collaboration systematic reviews (Dickson et al. 2000; Taylor-Robinson et al. 2007; and Taylor-Robinson et al, 2012). The most recently updated (2012) Cochrane review states clearly that, contrary to the policy recommendations of WHO and others, they find no convincing evidence of benefits of mass administration of deworming drugs, concluding that “it is probably misleading to judge contemporary deworming programs based on evidence of consistent benefit on nutrition, hemoglobin, school attendance, or school performance as there is simply insufficient reliable information to know whether this is so (2).” One of the authors of the review, in a subsequent interview, stated that he would “love [for deworming] to work. But to claim that it does on the basis of the evidence available is simply misleading” (Hawkes 2013). Another skeptical perspective comes from a comprehensive review by the charity rating organization Givewell, which concludes that “overall, evidence for the impact of deworming on short-term general health is thin.” With respect to longer-term developmental effects, they conclude that “empirical evidence...is very limited,” although they note that it is based on “two relatively well-known and well-executed studies” (referring to Bleakley 2007 and Baird et al 2011).⁵

1 The 2012 Cochrane review downgrades Alderman et al. (2007) because in in table 1 of the paper, the authors did not cluster standard errors for their comparison of treatment and control means, although they do so for their regression-based estimates.

² The Disease Control Priorities Project chapter on deworming (Hotez et al 2006) for example, presented a stylized case in which disability adjusted life years associated with STH could be prevented for just \$3.41, which would make deworming by far one of the most cost effective interventions available. However, researchers at Givewell demonstrated that these calculations were erroneous; see <http://blog.givewell.org/2011/09/29/errors-in-dcp2-cost-effectiveness-estimate-for-deworming/>.

³ See <http://www.copenhagenconsensus.com/copenhagen-consensus-iii/outcome>

⁴ See <http://www.povertyactionlab.org/policy-lessons/education/student-participation>

⁵ See <http://www.givewell.org/international/technical/programs/deworming>

A long back and forth has ensued between advocates of mass deworming and the authors of the Cochrane review. In response to the 2007 version of the Cochrane systematic review, Bundy et al (2009) criticized the Cochrane findings for not taking sufficient account of clustering (by overweighting studies that had individual-level randomization and by not including recent cluster randomized trials such as Miguel and Kremer 2004);⁶ for insufficient attention to data quality with respect to school attendance and cognitive outcomes, and sample attrition; and for not taking into account observational econometric studies with strong strategies for causal identification, such as Bleakley (2007).

The updated Cochrane review (2012) responded to these criticisms with several changes, including by incorporating the Miguel and Kremer study (but ranking it as relatively weak evidence based on the Cochrane evidence ranking criteria)⁷ and by adding several additional trials to the meta-analysis.⁸ In response to the negative findings about mass deworming in this updated Cochrane review, a group of economists, including Miguel and Kremer and other researchers associated with the Abdul Latif Jameel Poverty Action Lab (JPAL), and the NGOs Innovations for Poverty Action (IPA) and Evidence Action, responded in turn, disagreeing strongly with Taylor-Robinson et al. over the interpretation of Miguel and Kremer (2004) and over the non-inclusion of studies that use experimental or quasi-experimental variation generated by the original Miguel and Kremer study (such as Baird et al 2011 and Ozier 2012) to identify impact.⁹

This debate has led some to criticize the NGO Evidence Action (which recently incorporated the deworming-focused NGO Deworm the World) for scaling up school based deworming in the face of this contrary evidence provided in the Cochrane review. For example, Waddingham and Leach (2014) wrote on the International Initiative for Impact Evaluation (3ie) website that “On balance, the evidence does not favor the scaling up of [deworming].”¹⁰ The British medical Journal published an article in 2013 with the headline “Deworming Debunked.” (Hawkes, 2013).

This paper does not attempt to adjudicate this debate, but rather to add to the evidence base. One way to increase the evidence base would be to conduct new large scale randomized trials. In the meantime, however, policymakers have to make decisions in the face of competing

⁶ Because treatment has large potential spillovers to control groups by reducing community-level worm loads, they argue that individually randomized studies likely systematically underestimate the benefits of deworming. Therefore, they argue, “the primary focus of a review should be studies that use a cluster design.”

⁷ Cochrane reviews incorporate six potential sources of bias: random sequence generation bias, allocation concealment/selection bias, blinding (performance bias and detection bias); incomplete outcome/attrition bias; selective reporting bias, and “other” bias.

⁸ This study was apparently excluded from the 2007 review because of misunderstanding about the extent to which the results were robust when the sub-sample of communities also treated for schistosomiasis was excluded.

⁹ See <http://blogs.berkeley.edu/2012/07/20/cochrane-incomplete-and-misleading-summary-of-the-evidence-on-deworming/>

¹⁰ See <http://blogs.3ieimpact.org/how-much-evidence-is-enough-for-action/>

evidence. This points towards the value of using already-collected data, when possible, to answer questions about the impact of deworming programs. The Uwezo data sets (described further below) are particularly helpful in this regard. They are large scale, high-resolution education surveys conducted in areas in East Africa where important deworming programs have been implemented. They also represent, in effect, a blinded data collection exercise with respect to deworming, since (in the Uganda case, for example) the Uwezo team had no idea that they were sampling and surveying communities that had been part of a deworming project 7-10 years prior.

The deworming debate is complex and there is a wide range of potential outcomes to examine, which also vary in relevance according to the age of treated children. This paper addresses one component of the broader question about the comprehensive effects of deworming: it examines the effect of deworming pre-school age children on their test scores later in life. The channel through which deworming at young ages could affect young adult cognitive function is thought to be through improved nutrition and growth during the critical period of early growth and brain development. As mentioned above, the Cochrane Collaborative meta-analysis finds no convincing evidence that cognitive gains or school performance improvements result from deworming. However, relatively few of the studies examined by the systematic review are both cluster randomized and have long run cognitive outcomes. This suggests that the evidence is not yet definitive with regard to this outcome. Furthermore, one recent study (Ozier 2011) provides strong evidence for relatively large cognitive improvements among a cohort of Kenyan children whose older siblings were dewormed as part of the project described in Miguel and Kremer (2004). However, Ozier identifies this impact as operating through epidemiological spillovers rather than direct treatment of preschool age children. This paper seeks to add to the evidence base by examining the long run effects of a program that treated young children directly with deworming medication, rather than by reducing their exposure to worms through an epidemiological spillover mechanism.

3 The Uganda deworming project

The deworming project that is the focus of this paper took place in 48 parishes in 5 districts in Eastern Uganda from November 2000 to June 2003.¹¹ The districts were chosen because they were identified as having heavy worm loads, with at least 60% of children ages 5-10 infected, primarily with hookworm (Kabatereine et al 2001). It was implemented through the “Child Health Day” (CHD) delivery platform. Child Health Days are a pre-defined day in Uganda and other developing countries, usually every 6 months, in which all parents in a given catchment area are requested to bring all pre-school age children to a treatment site to receive a set of

¹¹ Parishes are the second lowest administrative level in Uganda, just above the village level and below the sub-county level.

basic health services such as Vitamin A supplementation, growth monitoring, and often any standard vaccines that they have not yet received. Five Child Health Days were held over the course of the project. The experimental variation was introduced in the following way: At Child Health Days in the control group, attendees were offered the standard intervention of Vitamin A supplementation, vaccines, growth monitoring, and complementary feeding demonstrations, while in the treatment group they were also offered deworming treatment, in the form of 400 mg of albendazole.¹² All children (except those who were ill at the time of the Child Health Day) between age 1 and 7 were offered albendazole in the treatment group, while none were offered it in the control group. The intervention was delivered by community-based organizations (CBOs). Randomization was at the parish level because that is the level at which CHDs are typically implemented in Uganda.¹³

3a Short run program evaluation results

Alderman et al. (2007) use data collected at the Child Health Day by program staff to measure anthropomorphic outcomes such as height and weight, and use baseline and end line surveys to measure population level participation.¹⁴ In regression models, they find that weight increases as a function of the number of CHDs attended, but by a larger increment in the treatment parishes than the control. For children with long gaps between treatment (corresponding roughly to annual treatment) weight gain was 5%, compared to 10% when treatment frequency was roughly once per year. The study population was comprised of children who had at least two anthropomorphic measurements, which means that the sample size was over 27,000 children in 48 clusters.

4 Data

Uwezo is a project led by the Tanzanian NGO Twaweza, modeled on India's Annual Status of Education Report (ASER). Uwezo does large-scale annual surveys to test basic literacy and numeracy in Kenya, Tanzania, and Uganda. The first test was done in Kenya in 2009 and in Tanzania and Uganda in 2010; for each country two years of data are in the public domain, and available at <http://www.uwezo.net>. Uwezo aims to collect data that is representative at district level, which means that in each year, 30 villages per district are sampled in each country, and 20 households per village are tested on basic literacy and numeracy. Over the course of its 2010

¹² The medicine (Zentel from GlaxoSmithKline) was provided in chewable form

¹³ According to the Uganda Service Provision Assessment Survey (2008), the catchment area of a parish-level health facility is up to 5,000 people.

¹⁴ Since only children who attended Child Health Days (CHDs) two times were measured, there is potential selection into treatment. However since both treatment and control were offered the standard CHD, differential selection would have to be based on the inclusion of deworming in the Child Health Day. This is possible, if we assume that parents know their childrens' worm infection status and are more likely to bring them to CHDs when they learn that deworming treatment is available.

and 2011 surveys in Uganda, Uwezo sampled 22 out of the 48 parishes that had participated in the deworming study in 2000-2003, surveying 1,097 children between the ages of 6 and 16. 763 out of these 1,097 children surveyed by Uwezo in 2010 or 2011 were between the ages of 1 and 7 during the deworming study period (2000 to 2003), and were therefore eligible to be dewormed.¹⁵ 10 of the parishes sampled by Uwezo were treatment parishes, and 12 were control. Table 1 (appendix) shows the relationship between age at survey and the age that the respondent had attained during the period of program implementation, from 2000-2003. Table 2 shows the means across a range of socioeconomic variables between treatment and control parishes, demonstrating that on observed socioeconomic variables in the Uwezo data there are no significant differences between respondents in the treatment and control parishes in 2010-2011.

Table 1: comparison of means, treatment and control parishes, combined 2010 and 2011 data

Variable	Obs	Treated	Control	T-C	p value
phone	1094	0.52	0.54	-0.03	0.55
radio	1093	0.58	0.65	-0.07	0.35
tv	1093	0.05	0.05	0.00	0.81
water	1093	0.10	0.00	0.10	0.25
electricity	1093	0.01	0.00	0.00	0.48
female	1094	0.49	0.51	-0.03	0.37
age	1094	10.77	10.53	0.24	0.23
mother age	1012	38.73	39.12	-0.40	0.82
female head	1094	0.16	0.16	-0.01	0.89
Mother post-primary	1012	0.10	0.11	-0.01	0.78
private	1041	0.01	0.00	0.01	0.30

**standard errors clustered at parish level*

5 Analysis

Given the random allocation of treatment, and the balance on observables between treatment and control parishes, I use a simple econometric framework to estimate the impact of deworming treatment. As shown in table 2, I estimate, in a regression framework, the effect of being a child in a deworming (“treatment”) parish during the 2000-2003 period. This means that I include all children in Uwezo’s sample who were aged 1-7 during the study period, and who are tested in 2010 or 2011 in one of the treatment or control parishes (710 children in 22 parishes). Tests scores are standardized, with mean zero and standard deviation of 1. The main result is that I observe sizable treatment effect from being dewormed, with treatment coefficients varying from 0.16 to 0.36 standard deviations. The coefficient is positive in all specifications, and it significant at the 5% level for math (with controls), at 10% level for English

¹⁵ Note that no control parishes were sampled by Uwezo in 2010.

(with controls) and math (without controls) and at 5% for the combined math/English score (with controls). In table 2, odd number specifications are the simple difference of means (using only the treatment indicator) while even numbered columns include controls for age, gender, survey round, and all interactions of age, gender, and survey year.¹⁶

Table 2: effect of deworming on test scores, all exposed cohorts

	(1) Math	(2) Math	(3) English	(4) English	(5) Total	(6) Total
Deworm parish	0.3009*	0.3620**	0.1640	0.2455*	0.2452	0.3181**
	(0.1545)	(0.1298)	(0.1707)	(0.1289)	(0.1707)	(0.1344)
<i>N</i>	710	710	710	710	708	708
<i>r</i> ²	0.0256	0.3425	0.0068	0.3486	0.0164	0.3878

Standard errors in second row, robust standard errors clustered at parish level. Even numbered columns contain (unreported) controls for age, gender, survey year, and interactions of all of these variables.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

I also test treatment intensity, or whether the effect size increases with the number of years exposed to treatment. This can be tested separately from age at exposure since different age cohorts got either zero, one, two, three, or four years of deworming depending on their age at program initiation and ending. I group respondents into having either received 0-1 year of treatment, or receiving 2-4 years of treatment. Effects are, in most specifications, two to three times larger in magnitude for the more intensely treated group; and I can reject that the effects are of equal magnitude for math scores, although not for the other two categories.

Table 3: effect of treatment intensity on effect size

	(1) Math	(2) Math	(3) English	(4) English	(5) Total	(6) Total
Treated 0-1 year	0.1592*	0.1272	0.2226***	0.1272	0.1872**	0.1655*
	(0.0905)	(0.0974)	(0.0675)	(0.0974)	(0.0757)	(0.0811)
Treated 2-4 years	0.3876**	0.3610**	0.2553	0.3610**	0.3382**	0.3129**
	(0.1431)	(0.1298)	(0.1529)	(0.1298)	(0.1554)	(0.1354)
<i>N</i>	1003	1003	999	1003	997	997
<i>r</i> ²	0.3393	0.5113	0.2692	0.5113	0.3324	0.5419

Standard errors in second row, robust standard errors clustered at parish level. Even numbered columns contain (unreported) controls for age, gender, survey year, and interactions of all of these variables.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

¹⁶ For comparability I follow the specification used by the other paper in this literature that attempts to measure long run cognitive impacts of early childhood deworming, i.e. Ozier (2011).

6 Treatment heterogeneity

In this section I test for two potential sources of treatment heterogeneity: gender, and wealth quintile. The first treatment interaction is by gender. Coefficients on the interaction term are positive in all specifications. The point estimates suggest an added benefit of deworming for female children; and the difference between the coefficient for male and the female interaction term is significant at 5% level for math (with controls) and at 10% for total score (with controls).

Table 4: treatment and gender interaction

	(1) Math	(2) Math	(3) English	(4) English	(5) Total	(6) Total
Deworm parish	0.1229 (0.1461)	0.2687** (0.1123)	0.0054 (0.1806)	0.1822 (0.1381)	0.0660 (0.1674)	0.2349* (0.1265)
Deworm*female	0.3525* (0.1721)	0.1814 (0.1390)	0.3207* (0.1608)	0.1328 (0.1202)	0.3587** (0.1663)	0.1687 (0.1231)
female	-0.1786 (0.1520)	-0.0802 (0.1234)	-0.0842 (0.1493)	0.0179 (0.1112)	-0.1317 (0.1514)	-0.0251 (0.1133)
<i>N</i>	710	710	710	710	708	708
r ²	0.0344	0.3263	0.0150	0.3261	0.0260	0.3701

Standard errors in second row, robust standard errors clustered at parish level. Odd columns include treatment indicator, female, and interaction. Even numbered columns contain (unreported) controls for age, gender, and survey year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Second, I test whether there are interactions with household wealth, as proxied by a household asset wealth index. Respondents are coded as belonging to a “poor” household if they fall into the poorest or second-poorest wealth quintile. Coefficients on the interaction of treatment and the poverty indicator are uniformly positive, suggesting an added positive effect for the poorest households. Equality of coefficient and interaction can be rejected at 5% significance for math (with controls), and at 10% significance for English and total score (both with controls).

Table 5: treatment and lower two wealth quintile interaction

	(1) Math	(2) Math	(3) English	(4) English	(5) Total	(6) Total
Deworm parish	0.2340 (0.1647)	0.2946** (0.1339)	0.0661 (0.1923)	0.1550 (0.1543)	0.1511 (0.1838)	0.2279 (0.1465)
Deworm*poor	0.1571 (0.1172)	0.1400 (0.1019)	0.2139* (0.1217)	0.1852 (0.1211)	0.2116* (0.1155)	0.1885 (0.1105)
Poor	-0.0252 (0.1131)	-0.1151 (0.0895)	-0.1477 (0.0926)	-0.244*** (0.0794)	-0.0976 (0.1046)	-0.1979** (0.0803)
<i>N</i>	710	710	710	710	708	708
r ²	0.0282	0.3258	0.0101	0.3307	0.0195	0.3723

Standard errors in second row, robust standard errors clustered at parish level. Odd columns include treatment indicator, poor indicator, and interaction. Even numbered columns contain controls for age, gender, survey year.

7 Robustness checks

In this section, I present a series of robustness checks. First, I examine the data for potential covariate imbalance (even though table 1 shows no statistically significant differences) and re-estimate our preferred specifications using only observations that have equal values of the baseline value. For example, although none of the components of the asset index (ownership of television, mobile phone, and radio, access to electricity, access to water) show statistically significant differences between treatment and control, the one variable from this index that shows any substantive (although non-significant) difference between treatment and control is access to water, which is almost 10 percentage points higher in the treatment group. Therefore to test if this imbalance drives the result, I restrict the sample to households without access to water. The results are robust to this adjustment. Another variable with some (though not statistically significant) baseline imbalance is the percentage of mothers with no education, which is higher in the control group. I again re-estimate the main specifications among mothers with at least some education. The results do not change. Second, I include district fixed effects. In this specification, we can reject the null at $p=0.05$ for all five out of the six regressions. Third, I generate a new, “placebo” treatment group from among the list of other parishes in the study area which were sampled by Uwezo but which were not chosen as participants in the deworming study. In this model the “treatment” variable is positive but the effect does not reach statistical significance, either in the main specifications or when socioeconomic variables are used as controls. I also estimate regressions using data from only the 2011 Uwezo survey round, since no control parishes were selected in 2010, and I (separately) exclude the two parishes that were initially selected for the study but did not participate. Finally, I re-estimate the main specifications with controls for wealth quintiles, generated using the first principal component of an asset index. The main results come through in each of these specifications, and in most instances significance is strengthened. Each of these regression tables is presented in the appendix (tables 10-13).

Another potential concern is that, since Uwezo only randomly sampled 22 out of the 48 study clusters from the original experiment, normal regression estimates with clustered standard errors might be overly likely to reject the null hypothesis, unless an adjustment for small number of clusters is applied. Therefore table 7 re-estimates the original specifications using Wild cluster bootstrapped standard errors. In these specifications the results for math are still highly significant while the English and the total score coefficients are significant at the 10% level when district fixed effects are included.

Table 6: main model with Wild cluster bootstrapped (Cameron et al. 2008) standard errors

	(1) Math	(2) Math	(3) English	(4) English	(5) Total	(6) Total
Deworm parish	0.281**	0.303*	0.154	0.193*	0.226	0.26*
p values	0.04	0.09	0.3	0.1	0.13	0.1
R-squared	0.303	0.342	0.29	0.337	0.338	0.39
N	715	715	714	714	712	712

p values in second row, robust standard errors clustered at parish level. Odd-numbered columns have gender and age controls, even numbered columns have gender and age controls and district fixed effects.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

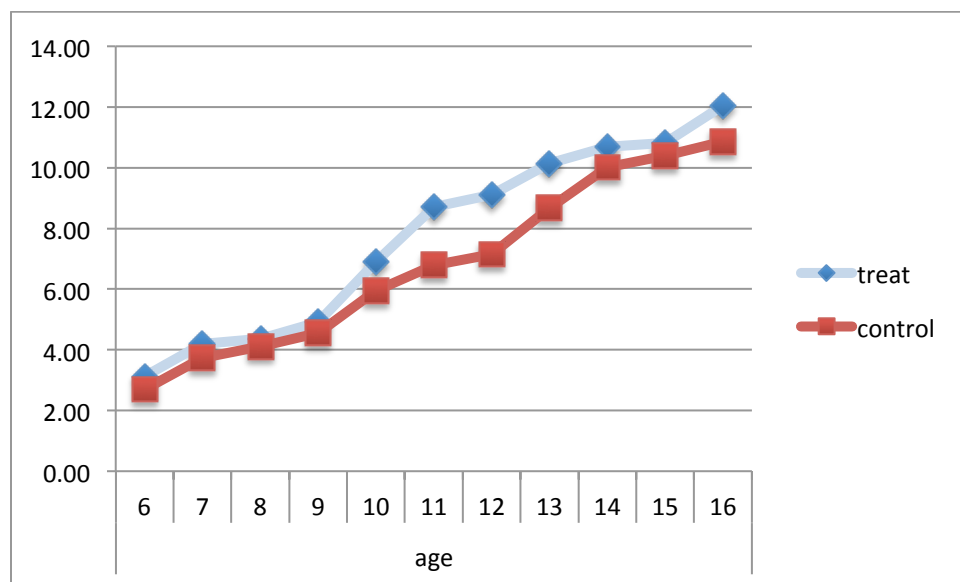
I also present p-values derived non-parametrically, using randomization inference hypothesis testing methods. Here the procedure is as follows. First, I collapse the values into parish level averages, leaving 22 values. Then I derive a distribution of potential outcomes by simulating a world in which new randomly generated combinations of treatment and control, using these 22 parishes, is run. I then rank the coefficients and see whether the true observed coefficients fall within the 95% confidence interval generated by this procedure. In table 8 I report the p values for 100 iterations of this test for each outcome. In this test, math scores are still significant at $p=0.4$ while English is no longer significant at conventional levels ($p=0.13$), while the total score is marginally significant at $p=0.08$. (See also appendix figures 1-3).

Table 7: Randomization inference hypothesis test

	Math	English	Total
Treatment			
Number of permutations	100	100	100
observations	22	22	22
P values	0.04	0.13	0.08

A final, intuitive robustness check is to simply examine the unadjusted pattern of scores by age. The youngest cohorts surveyed in the treatment group were too young to benefit from multiple years of the program (for example, child aged 7 in 2010 were born in 2003 and could only have been dewormed twice at most), and the oldest children were too old (a 16 year old in 2003 would have aged out of the 1-7 age group after 1 year of the program), it seems reasonable to expect small differences between treatment and control among the *oldest groups* and the *youngest groups*, and larger differences in the middle age groups. That is exactly the pattern that can be seen in the unadjusted test scores, as figure 1 demonstrates.

Figure 1: total scores by age



8 Discussion

The finding that early childhood deworming appears to have long lasting positive impact on test scores, and (implicitly) cognitive ability is, on one hand, very consistent with Ozier’s (2011) finding that children in Western Kenya who were dewormed under 1 show large cognitive benefits compared to children dewormed at older ages, and the long run labor market gains shown by Baird et al (2011). And they are consistent with the significant weight gain that Alderman et al’s (2007) regression models suggest was the effect of the deworming program in the short run. On the other hand, they are in tension with the findings of the Cochrane Collaborative systematic review, which finds no evidence of weight gain or cognitive benefit. How can these findings be reconciled? While the section 7 examined the statistical robustness of the estimates presented in this paper, in this section I take a step back to consider the plausibility of the effect sizes identified, potential mechanisms, and consistency with the broader literature.

8a Plausibility of effect size

One circumstance under which these estimates might not be plausible would be if the original deworming had only reached a small fraction of the treatment population. Since I measure what are in effect “intent to treat” estimates, and yet still see large effects, this implies that the effects are actually proportionally larger for those children actually dewormed (we do not observe actual deworming status in the Uwezo survey; all we know is whether children live in parishes which were provided with deworming). For example, if on average only 25% of children were ever dewormed by the program in treatment parishes, then Uwezo’s random

sample would, on average, be comprised of 75% untreated children and 25% “ever treated” children. Assuming no spillovers (such that only treated respondents benefited from deworming), large observed effects at population level would imply very large, perhaps implausible, “treatment on treated” (TOT) effects – on the order of 0.8-1.2 standard deviations.

But program coverage, as measured by Alderman et al (2007) was actually quite high. 66% of children between ages 1-7 in treatment districts report being dewormed in the past two years, and 74% reported attending at least one child health day in the past two years. The average child attended 1.74 child health days over the three year program period.¹⁷ Thus, depending on whether one uses the 66% figure or the 74% figure as closer approximation of population coverage, the proportion of children in study parishes (and in the sample) that were dewormed at least once is therefore likely between two-thirds and three quarters. These implied TOT effects (with the conservative assumption of no spillovers), are therefore large but not so large as to be ex ante implausible. Another factor working against the plausibility of the estimates presented here it that there was also apparently some crossover or contamination between treatment and control. While coverage of deworming tripled between 2000 and 2003 (from 22% to 66%) in treatment parishes, it also reportedly increases by approximately 50% in the control parishes (from 24% to 35%).¹⁸ When we scale our implied TOT effects to account for this crossover, we are again at an effect size close to 1 standard deviation. But again, this is under the conservative assumption of no spillover effects. Yet in a spatially and temporally nearby setting (the other side of the Kenyan border, 3-5 years earlier), both Miguel and Kremer (2004) and Ozier (2011) document sizable spillovers from deworming treatment. Without knowing more about the magnitude of the spillover in the Ugandan case, it is difficult to precisely judge the plausibility of the implied effect size.

8b Potential mechanisms

Although the fact that identification in this study is based on the original randomization should generate confidence in the internal validity of these results, we unfortunately lack corroborating information that could suggest the precise mechanism might be driving this result. For example, if the early life weight gain induced by randomization translated into greater observed adolescent height in the treatment group, we could be confident that we are observing a biomedical pathway linked to early life nutrition (via deworming), and not a statistical artifact. Similarly, we would ideally have a broad range of cognitive tests to better identify the channel through which improved numeracy and literacy is manifested. Unfortunately, the flip side of Uwezo’s extremely large sample is that the survey instrument itself is quite limited. This means that we do not have any evidence on health outcomes from

¹⁷ See Alderman et al. Table 4.

¹⁸ In both groups, the average child attended 1.74 child health days over the three year period.

Uwezo, such as anthropometry, or detailed cognitive measurements beyond basic numeracy and literacy. As a result, we cannot test whether specific health channels are operational, or which components of cognition are driving improved math and English scores. Nor, since we do not have information on program attendance at the original deworming program, can we determine whether these effects are driven by the respondents who were actually dewormed, or whether spillover effects play a large role. Another limitation is that this study did not compare a deworming program versus a pure control, but rather deworming plus a standard Child Health Day (Vitamin A, growth monitoring) versus a standard Child Health Day without deworming. Therefore if there are any interactions between deworming and other components of the Child Health Day package, the results from this context should be interpreted in light of this. Finally, it is also the case that shortly after the trial, the Ugandan Ministry of Health made deworming a standard component of Child Health Days. Thus, after the study period, both groups had access to routine, free deworming treatment.

Retuning to the question of causal mechanisms, the one channel that can be tested relates to school enrollment. If there were persistent *health* effects of early childhood deworming (as distinct from cognitive or growth effects) we might see that children dewormed in early life attend school more (in 2009/2010) or are less likely to never have enrolled in school between when they were dewormed in 2000-2003 and when they were surveyed in 2009/2010. However there is no relationship between these measures and treatment status.

Table 8: school enrollment as a function of deworming

	(1)	(2)	(3)	(4)
	In school	In school	Never enrolled	Never enrolled
Dewormed parish	0.0186	0.0157	0.0071	0.0060
	0.0129	0.0137	0.0070	0.0070
<i>N</i>	763	763	704	704
<i>r</i> ²	0.0027	0.0277	0.0015	0.0163

Standard errors in second row, robust standard errors clustered at parish level. Even numbered columns contain (unreported) controls for age, gender, survey year, and interactions of all of these variables.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

8c Consistency with the broader literature

Soil transmitted helminths are an infectious disease, so an individual's likelihood of infection decreases by some unknown magnitude as his neighbors' levels of infection decrease, making spillover effects a potentially important factor. However, the vast majority of the evidence used in systematic reviews (and driving the systematic review conclusion that mass deworming is ineffective) comes from studies randomized at the individual level (and therefore with results

potentially attenuated by spillovers). For example in the second Cochrane review (Taylor-Robinson et al 2007), only 3 out of 34 included trials were cluster randomized and analyzed with design effects¹⁹, while in the 2012 updated review, 8 out of 42 were cluster randomized, and only 5 of these were originally analyzed using design effects (in 2012 the review team re-analyzed using imputed design effects if results were not originally reported in this way.)

The three cluster randomized trials in the 2007 Cochrane review are Stoltzfus et al (1997); and two trials led by Awasthi et al. (2000; 20001). These studies, unlike many of the individually randomized ones, find positive treatment effects. Stoltzfus et al. found positive effects on growth among a sub-group (children under 10) school aged children on the Zanzibari island of Pemba in a school-based deworming program after 1 year, using a single dose of mebendazole. Awasthi et al (2001) studied twice-annual doses of albendazole in 60 urban clusters of Lucknow, in Uttar Pradesh state in India. After 1.5 years treatment group showed significant gains in weight.²⁰ Awasthi et al. (2000) examined albendazole treatment in the same slums (in a separate trial) and found reduced risk of stunting in the treatment group. The 2012 Cochrane review added in Miguel and Kremer (2004) and Alderman et al. (2007), both of which were cluster randomized trials with positive treatment effects. It also added in Rousham (1994) and Hall (2006), which were cluster randomized trials without positive and significant effects. Nonetheless it is suggestive that while the overall Cochrane meta-analysis shows no effect, the majority of clustered trials show positive effects. This is consistent with the critique offered by Bundy et al (2009).

However, since the publication of the 2012 updated systematic review, new research has come to light which further complicates any simple narrative about clustered versus unclustered trials. In 2013 the results of the “DEVTA” trial from northern India, the largest deworming trial to date, were published. The DEVTA trial (which was a factorial design trial of Vitamin A supplementation and deworming) showed no significant impact of deworming on children’s weight, in a large, cluster-randomized sample in Uttar Pradesh. (The difference in weight between treatment and control was significant at the 10% level but small in magnitude, at 0.04 kg). While the DEVTA trial took place in an environment of relatively low worm prevalence, it is difficult to say whether this explains the differences between this outcome and that of other cluster-randomized trials. Further trials in high prevalence settings, and with specific focus on pre-school age children, seem justified at this stage.

9 Conclusion

¹⁹ Six of these trials were cluster randomized, but three of these did not use design effects in analysis and so are not included in the meta-analysis.

²⁰ The 2012 review includes these three studies plus Miguel and Kremer (2004); Alderman et al (2007); Hall 2006; and Rousham (1994).

Mass deworming of school age children is a highly touted policy by a range of authoritative sources – and mass deworming of preschool age children has seemed to be a promising extension of this policy. However, the evidence base has been clouded by the current controversy over the effect of deworming, as shown by the disagreements between the researchers associated with the Cochrane Collaborative review, and the group of deworming proponents, associated with IPA, JPAL, Evidence Action, the World Bank, and other institutions. This paper exploits a new data source to identify large educational benefits to a group of school children dewormed in early childhood - effects which are present despite the fact that educational outcomes are measured 7-10 years after the end of the deworming experiment. It avoids weaknesses in previous studies by exploiting a cluster-randomized approach (thus avoiding attenuation of effect via spillovers). As such, it strengthens the case that there are important and persistent cognitive benefits to mass deworming in settings of high worm prevalence.

References

Alderman, H, et al. (2007). Alderman, H., J. Konde-Lule, I. Sebuliba, D. Bundy, and A. Hall (2006): "Effect on weight gain of routinely giving albendazole to preschool children during child health days in Uganda: cluster randomised controlled trial," *British Medical Journal*, 333, 122–127.

Awasthi S, Pande VK. Six-monthly de-worming in infants to study effects on growth. *Indian Journal of Pediatrics* 2001;**68**(9):823–7.

Awasthi S, Pande VK, Fletcher RS. Effectiveness and cost- effectiveness of albendazole in improving nutritional status of pre-school children in urban slums. *Indian Pediatrics* 2000;**37**(1):19–29.

Baird S, Hicks JH, Kremer M, Miguel E. Worms at work: long-run impacts of child health gains. 2012. Accessed at <http://scholar.harvard.edu/kremer/publications/worms-work-long-run-impacts-child-health-gains>.

Bleakley, Hoyt. (2007). "Disease and Development: Evidence from Hookworm Eradication in the American South." *Quarterly Journal of Economics*, 122(1):73-117.

Bundy, D. A. P., M. Kremer, H. Bleakley, M. C. H. Jukes, and E. Miguel (2009): "Deworming and Development: Asking the Right Questions, Asking the Questions Right," *Public Library of Science: Neglected Tropical Diseases*, 3(1), e362.

Gerber A, Green D. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: WW Norton.

Hall A, Nguyen Bao Khanh L, Bundy D, Quan Dung N, Hong Son T, Lansdown R. A randomized trial of six monthly deworming on the growth and educational achievements of Vietnamese school children. Unpublished manuscript.

Hawkes, N. 2013. "Deworming Debunked." *British Medical Journal*. Accessed at <http://www.bmj.com/content/bmj/346/bmj.e8558.full.pdf>. doi: 10.1136/bmj.e8558

Hotez P et al. "Helminth Infections: Soil Transmitted Helminth Infections and Schistosomiasis." In *Disease Control Priorities in Developing Countries*. 2006. Jamison, DT et al (eds). Washington DC: World Bank, and New York: Oxford University Press.

Kabatereine, N. B., Tukahebwa, E., Brooker, S., Alderman, H., & Hall, A. (2001). Epidemiology of intestinal helminth infections among schoolchildren in Southern Uganda. *East African Medical Journal*. 78(6), 283–286.

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

Ministry of Health (Uganda) and Macro International (2008). Uganda Service Provision Assessment 2007. Kampala, Uganda: Ministry of Health, and Calverton, MD: Macro International.

Ozier, O. (2010). "Exploiting Externalities to Estimate the Long-Term Benefits of Early Childhood Deworming", unpublished working paper, U.C. Berkeley.

Northrop-Clewes CA, Rousham EK, Mascie-Taylor CN, Lunn PG. Anthelmintic treatment of rural Bangladeshi children: effect on host physiology, growth, and biochemical status. *American Journal of Clinical Nutrition* 2001;**73**(1): 53–60.

Stoltzfus RJ, Albonico M, Tielsch JM, Chwaya HM, Savioli L. School-based deworming program yields small improvement in growth of Zanzibari school children after one year. *Journal of Nutrition* 1997;**127**(11):2187–93.

Taylor-Robinson D, Jones A, Garner P. (2007) "Deworming drugs for treating soil-transmitted intestinal worms in children: effects on growth and school performance." Cochrane Database of Systematic Reviews 2007, 4, CD000371.doi:10.1002/14651858.CD000371.pub31.

Taylor-Robinson DC, Maayan N, Soares-Weiser K, Donegan S, Garner P. Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin and school performance. *Cochrane Database Syst Rev* 2012;11:CD000371.

Waddington, H and Leach, B. March 4, 2014. "How Much Evidence is Enough for Action?" Blog post accessed at <http://blogs.3ieimpact.org/how-much-evidence-is-enough-for-action/>.

Table 9: age at survey / age at treatment for 2010 and 2011 survey rounds

survey year age in year:	2011 2000	2010	2011 2001	2010	2011 2002	2010	2011 2003	2010
6	-5	-4	-4	-3	-3	-2	-2	-1
7	-4	-3	-3	-2	-2	-1	-1	0
8	-3	-2	-2	-1	-1	0	0	1
9	-2	-1	-1	0	0	1	1	2
10	-1	0	0	1	1	2	2	3
11	0	1	1	2	2	3	3	4
12	1	2	2	3	3	4	4	5
13	2	3	3	4	4	5	5	6
14	3	4	4	5	5	6	6	7
15	4	5	5	6	6	7	7	8
16	5	6	6	7	7	8	8	9

Table 10: placebo treatment test

	(1) Math	(2) Math	(3) English	(4) English	(5) Total	(6) Total
Placebo treat	0.2591	0.2242	0.1604	0.0877	0.2392	0.1966
	(0.1623)	(0.1768)	(0.1715)	(0.1803)	(0.1697)	(0.1844)
<i>N</i>	751	751	749	749	747	747
<i>r</i> ²	0.0188	0.2458	0.0064	0.2435	0.0151	0.2852

Standard errors in second row, robust standard errors clustered at parish level. Even numbered columns contain (unreported) controls for age, gender, survey year, and interactions of all of these variables.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11a: analysis restricted to households without water supply at home (odd columns) or mothers without education (even columns)

	(1) Math	(2) Math	(3) English	(4) English	(5) Total	(6) Total
Deworm parish	0.3903***	0.1951**	0.2757**	0.1362	0.3550**	0.1951**
	0.1277	0.0843	0.1323	0.1262	0.1319	0.0843
<i>N</i>	670	510	670	510	668	510
r2	0.3480	0.3476	0.3543	0.3345	0.3945	0.3476

Standard errors in second row, robust standard errors clustered at parish level. All columns include age, gender and survey year indicators, and all interactions off these variables.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11b: Full sample, controlling for lack of water supply/mother's education

	(1) Math	(2) math	(3) English	(4) English	(5) Total	(6) Total
Deworm parish	0.3911***	0.2848***	0.2762*	0.1938	0.3561**	0.2848***
	0.1279	0.0891	0.1328	0.1177	0.1325	0.0891
Water	-0.2249***		-0.2373**		-0.2914***	
	0.0707		0.0923		0.0596	
Mother no education		-0.4222**		-0.2922**		-0.4222**
		0.1901		0.1316		0.1901
<i>N</i>	710	653	710	653	708	653
r2	0.3452	0.3500	0.3514	0.3340	0.3923	0.3500

Standard errors in second row, robust standard errors clustered at parish level. All columns include age, gender and survey year indicators, and all interactions off these variables. Odd columns control for whether the household has water access, even columns control for whether the mother has any education.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Additional socioeconomic controls (wealth quintiles, private school attendance)

	(1)	(2)	(3)	(4)	(5)	(6)
	Math	Math	English	English	Total	Total
Deworm parish	0.3696*** (0.1270)	0.2877*** (0.0900)	0.2523* (0.1323)	0.1908 (0.1236)	0.3265** (0.1331)	0.2534** (0.1086)
N	710	652	710	650	708	650
r2	0.3526	0.3613	0.3609	0.3502	0.4012	0.4045

Standard errors in second row, robust standard errors clustered at parish level

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: district fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Math	Math	English	English	Total	Total
Deworm parish	0.301** (0.126)	0.328** (0.129)	0.183 (0.107)	0.228** (0.092)	0.253** (0.116)	0.292** (0.111)
N	715	715	714	714	712	712
r2	0.078	0.342	0.07	0.344	0.086	0.394

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

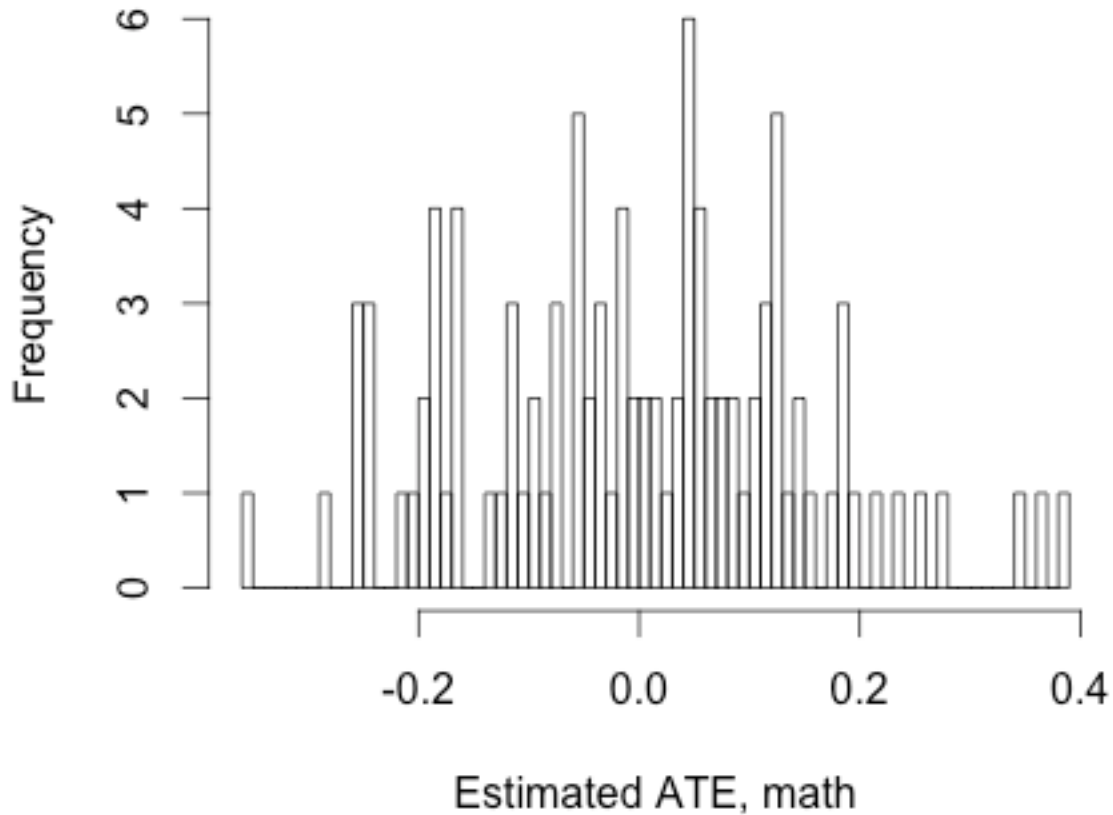
Standard errors in second row, robust standard errors clustered at parish level

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

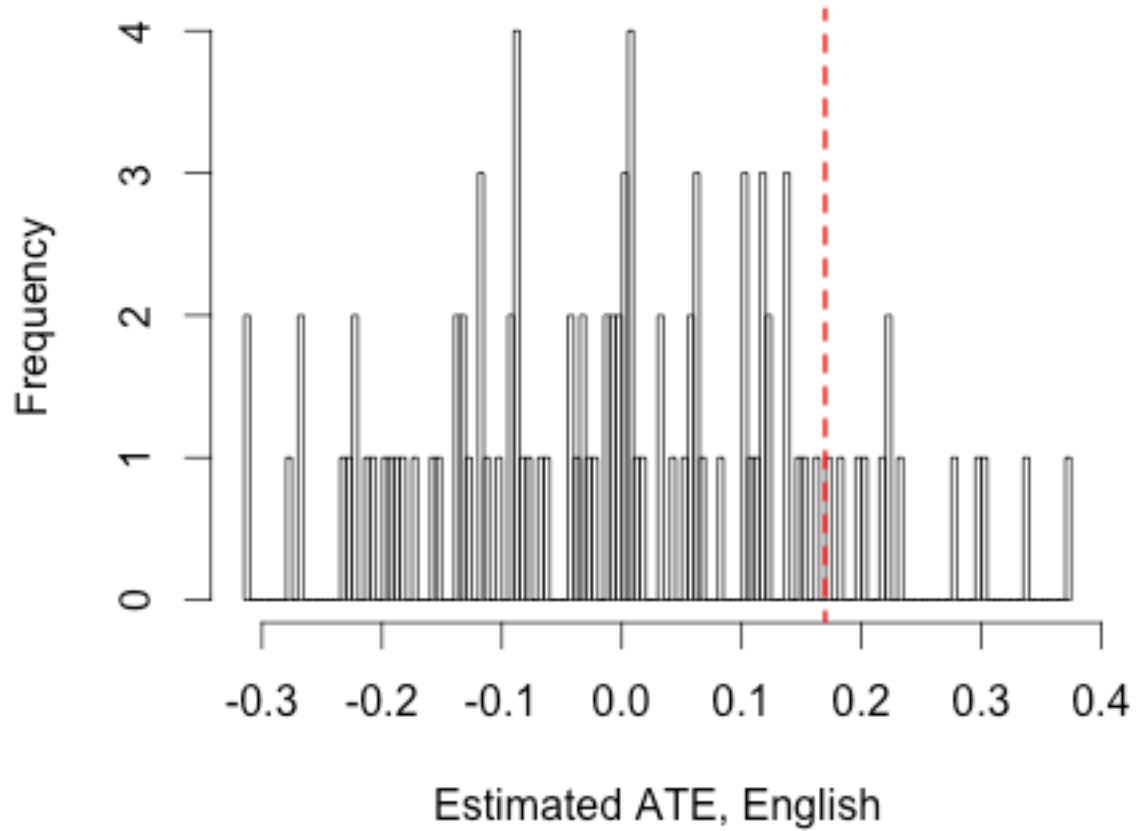
Appendix: Randomization inference figures

Tables generated using R code adapted from Gerber and Green (2012)

Distribution of the Estimated ATE for math under the Sharp Null Hypothesis of No Treatment I $p = 0.04$, $SE = 0.152$



**Distribution of the ATE for English
under the Sharp Null Hyp of No Treat Effect
 $p = 0.13$, $SE = 0.153$**



**Distribution of the ATE for total score
under the Sharp Null Hyp of No Treat Effect
 $p = 0.08$, $SE = 0.158$**

