Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming

Owen Ozier*

World Bank - Development Research Group

July 27, 2016

Abstract

I investigate whether a school-based deworming intervention in Kenya had long-term effects on young children in the region. I exploit positive externalities from the program to estimate impacts on younger children who were not directly treated. Ten years after the intervention, I find large cognitive effects—comparable to between 0.5 and 0.8 years of schooling—for children who were less than one year old when their communities received mass deworming treatment. I find no effect on child height or stunting. Because treatment was administered through schools, I also estimate effects among children whose older siblings received treatment directly; in this subpopulation, effects on cognition are nearly twice as large. (*JEL:* I10, O12, O15)

*Research supported by: UC Berkeley (IBER, CEGA, Fernald Lab, CEG, John Carter Endowment, Rocca Fellowship, BBB, Henry Wheeler Center); USAID; PCD; WUSTL Olin Business School CRES. KEMRI and UCB CPHS conducted IRB review. Jamie McCasland and John Ikoluot provided superb field research assistance and team leadership. Thanks to Harold Alderman, Prashant Bharadwaj, Simon Brooker, Donald Bundy, Eric Edmonds, Willa Friedman, Lia Fernald, John Hoddinott, Gerald Ipapa, Pamela Jakiela, Kelly Jones, Matthew Jukes, Carol Kemunto, Rose Kimani, Joseph Konde-Lule, Michael Kremer, Karen Levy, Edward Miguel, Rohini Pande, Adam Wagstaff, Dean Yang, and numerous conference and seminar participants for suggestions throughout this work. First version: October 17, 2010. The findings, interpretations, and conclusions expressed in this paper are entirely those of the author, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. All errors are my own. Please direct correspondence to oozier@worldbank.org.

1 Introduction

Shocks in early childhood can permanently transform an individual's potential lifetime health, earnings, and cognition. Several variations of this idea, as hypothesis or as stylized fact, are well-known. The lasting effects of nutrition shocks a child experiences in early childhood are referred to as the "Barker Hypothesis;" specifically before birth, the "fetal origins" hypothesis (Almond and Currie 2011b). The terms "critical" and "sensitive" are applied to periods in a child's life during which inputs are most important, particularly for cognitive development (Cunha and Heckman 2008, Knudsen 2004). Yet because of the demanding data required, few studies have established causal relationships between public policies intervening early in childhood and long-term outcomes.

In this paper, I examine the treatment of a disease that, while rarely fatal, is highly prevalent among children around the world: intestinal parasites. These helminths (worms) infect more than one billion people worldwide: predominantly young children in Asia and Sub-Saharan Africa (Hotez, *et al.* 2006). Helminth infections directly cause anemia and listlessness, and may result in chronic symptoms (Bleakley 2007). Despite the potential health benefits of reduced worm infections, the large spillovers and short-run discomfort of deworming may make individuals unlikely to privately finance deworming, leaving public policy to coordinate their actions (Dupas 2011). Current research suggests that mass administration of deworming medication to school-age children may be one of the most cost-effective possible ways to increase school attendance and improve adult outcomes (Miguel and Kremer 2004; Bleakley 2007; Bundy, *et al.* 2009).

A recent review, however, examining school performance and cognition among an array of outcomes, found either insufficient reliable information on whether deworming treatment has any beneficial effect, or simply that there is no effect at all (Taylor-Robinson, et al., 2015). This, accompanied by debate over the most appropriate way to replicate the paper by Miguel and Kremer (2004), has lead to a broad discussion on the need for additional evidence on the impacts of deworming (Evans 2015).

In this paper, I present one of the first pieces of evidence on the long-term effects of reducing helminth infection in *early* childhood by exploiting externalities from a randomized deworming intervention in Kenya.[1] I take a novel approach to the phased randomized intervention first studied by Miguel and Kremer (2004), in which deworming was randomly introduced into schools (once initiated, the program remained present): unlike Miguel and Kremer (2004) and the follow-up study by Baird, Hicks, Kremer and Miguel (2014), I follow a different, younger cohort of respondents. Though mass deworming efforts are often aimed only at school-age children, these interventions have large epidemiological spillovers both on other schoolchildren (Miguel and Kremer 2004) and on others in the community (Bundy, *et al.* 1990). Taking advantage of these spillovers, I gathered data in 2009 and 2010 in order to compare children who were in their first years of life at the time that treatment started at their community's school to children from the same cohorts in the catchment areas of as-yet untreated schools. Because the intervention for schoolchildren in Kenya had such large spillover effects, I hypothesize that children who were not yet old enough to attend school also garnered benefits. Because of their age at the time of the intervention, I further hypothesize that these younger cohorts may have been more sensitive to the intervention than the older children who actually received the drugs. Until recently, however, pre-school-age children were neither the focus of deworming efforts, nor were they the subject of much deworming research, because their parasitic load is typically much lower than it is in older children.[2]

I find large effects on cognitive performance equivalent to half a year of schooling, robust to a variety of specifications, more than ten years after the original intervention. Effects are strongest among those whose older siblings were likely to have been in school at the time of the original intervention. This

[1] A working paper by Croke (2014) follows a population in which many young children directly received deworming treatment (as I discuss in greater detail in Section 2.4), while this study relies on spillovers from a (likely) higher-compliance intervention. As such, that project and the one here are complementary but distinct contributions.

[2] Stoltzfus, et al., (2004), p.368, write: "the main programmatic focus of helminth control has been school-age children."

is particularly true when those siblings are female, as one might expect given the frequent role of older sisters as caregivers, and the epidemiological channel that frequent physical contact might create. I do not find any statistically significant long-term effect on measures of stunting or height. This is not the first paper to find an effect of early childhood interventions on cognition without finding an effect on height, however: my results are consistent with the literature suggesting that, however sensitive to nutrition shocks physical development is, cognitive development is even more so (Lewit and Kerrebrock 1997).

This study has a limitation: I did not directly observe the subjects in this study during the first years of their lives (when the deworming took place), so I cannot directly measure the change in their worm infection rate brought about by the deworming of the older children living nearby. Nevertheless, I carry out a range of tests to address the possibilities of other mechanisms being at play. I consider threats to validity, including differential fertility or migration, and find evidence for neither. I consider the possibility that the intervention's spillovers operated through increased stimulation from siblings' (potentially) changed abilities, rather than the more obvious epidemiological channel. The test I am able to perform does not support this alternative channel.

Instead, my results support the theories that sensitive periods in early childhood are essential for cognitive development, and provide evidence that inexpensive actions are available that could produce lasting improvements in the lives of millions via health improvements early in life. The main result, along with the identification strategy, is summarized in Figure 1, and is discussed in greater detail in the sections that follow.

A recent widely-publicized review of deworming studies by Taylor-Robinson, et al. (2015), was able to include only two cluster-randomized studies of deworming that included cognitive outcomes. Only one is published (that of Miguel and Kremer, 2004), and neither found impacts. Critically, neither is a study of deworming in *early* childhood; the review is silent on the long-term cognitive effects of deworming in early childhood. My findings and those in the working paper by Croke (2014) corroborate one another in both filling

the research gap that the review makes salient and changing the balance of evidence: This study and Croke (2014) find large, lasting, positive effects of early childhood deworming on cognition.

The remainder of this paper is organized as follows: In Section 2, I discuss the literature on the nature of the disease, the original intervention in Kenya, and critical periods; Section 3 provides details on the new data collection undertaken in 2009 and 2010; Section 4 presents the identification strategy and main equations to be estimated; in Section 5, I discuss estimation results and a variety of robustness checks and alternative specifications in light of the existing literature; in Section 6, I weigh relative costs and benefits from the perspective of government expenditures and revenues; and Section 7 concludes.

2 Background

2.1 Biology of helminth infection: biology and policy

A handful of helminth species are responsible for infecting at least one billion people (Hotez, *et al.* 2006). This group comprises schistosomes, along with soil-transmitted "geohelminths:" roundworm, whipworm, and hookworm.[3] Several of these species are endemic in western Kenya, and though these infections can be addressed inexpensively with existing drugs, they usually go untreated.[4] All of these parasites inhabit parts of the human digestive tract; female worms produce eggs that spread via human excrement.[5] Subsequent infection of new hosts follows different routes depending on the parasite species. In the case of whipworm and roundworm, an individual is infected by ingesting a worm egg (often from contact with soil contaminated with feces). Other

[3]Relevant species are *Schistosoma mansoni* and *haematobium*; *Ascaris lumbricoides*; *Trichuris trichiura*; and hookworms *Necator americanus* and *Ancylostoma duodenale*.

[4]Albendazole and mebendazole are anti-geohelminth medications. Schistosomiasis is usually treated with praziquantel. Diagnosis, which involves laboratory work, is much more expensive than the medications themselves; mass deworming with these modern drugs only began in the 1990s, but has recently become more common.

[5]Here, I discuss *Schistosoma mansoni* rather than *Schistosoma haematobium*, as urinary schistosomiasis is not endemic in western Kenya.

species infect simply via human contact with worms in a particular phase of their life cycles.[6]

From a public policy perspective, the economics of diagnosis and treatment favor mass deworming over making treatment conditional on individual screening: diagnosis is expensive; treatment is cheap; and side-effects of treatment for the uninfected are minimal (Ahuja, et al., 2015). Thus far, school-age children have been emphasized in studies of deworming because they are known to host the highest numbers of parasites (Bundy 1988). However, very recent studies, reviewed by Albonico, *et al.* (2008), also document child health improvements in response to early childhood deworming. Despite promising short-term results, no published study to date has shown whether early childhood deworming can have lasting benefits.

The current debate on deworming reflects limitations of the current evidence base. Few studies have been designed with long-run health and education outcomes in mind, or with the epidemiological spillovers of deworming treatment taken into account. As such, a recent review, focusing on randomized trials, found insufficient evidence on the question of whether deworming affects school-related outcomes. Constrained to a handful of studies, it then interprets the current literature as finding no evidence for effects of deworming on cognition (Taylor-Robinson, et al., 2015).

There remains substantial room for additional research in this area, as recent discussion has pointed out (Bundy, Walson and Watkins 2013). To take an example pertinent to this study, in order to reach the conclusion that there is no evidence for effects on cognition, the recent review chose study inclusion criteria that forced it to rely upon just two studies with short follow-up periods and low baseline rates of worm infection.[7]

[6]Hookworms often penetrate the skin through the sole of the foot, while schistosomes enter the skin through lake or river water while part of a person's body is immersed (Bundy et al. 2001, Mott 2001).

[7]The criteria of the Taylor-Robinson, et al. review focused on randomized or quasi-randomized trials. Bleakley's 2007 historical analysis of a natural experiment, finding long-term cognitive effects of deworming in the United States, was thus excluded from consideration. Bleakley's study follows impacts of deworming into adulthood, using a measure of treatment varying from 0 to 19 years of childhood; it provides clear evidence of lasting

2.2 Past intervention

Between 1998 and 2001, Miguel and Kremer (2004) randomly phased in deworming drugs to a group of 75 primary schools in western Kenya, in the "Primary School Deworming Project," PSDP: once PSDP deworming started in a school, it was continued thereafter. Children in this region suffered from high rates of worm infection: at baseline, 92 percent of children had at least one type of worm infection, and many were infected with multiple species of worm; hookworm alone infected more than 70 percent of school-age children (Brooker, Miguel, Moulin, Luoba, Bundy and Kremer 2000, Miguel and Kremer 2004). The school-based deworming program therefore followed a mass-deworming protocol rather than individual testing before treatment. The program reduced infections and increased school attendance. Only schoolchildren were dewormed, but the authors found large spillovers within the community: in terms of school attendance, for example, children in dewormed areas who were not actually given medication still received around 60% of the benefits of direct deworming.[8] This is consistent with evidence from the island of Montserrat, where mass deworming of children aged 2-15 may have reduced parasitic loads in adults who received no medication (Bundy, *et al.* 1990). No significant test score gains were documented in the years immediately following the PSDP intervention, however. Thus far, the long-term effects of the intervention in Kenya on the recipients of medication have included, in early adulthood, clear increases in wage and hours worked (Baird et al. 2014). Improvements in labor market outcomes are attributed to improved health; test scores suggest increased human capital in the long run, but not through general intelligence

benefits of deworming, but not specific benefits in early childhood.

[8]Baird et al. (2014) summarize the spillover in Miguel and Kremer (2004) as being 78 percent as large as the direct effect of deworming. The spillovers I rely upon in this paper are the within-community spillovers that Miguel and Kremer observed within schools; I do not rely on the longer distance spillovers Miguel and Kremer found. The recent replication work of Aiken, Davey, Hargreaves and Hayes (2015) calls attention to a coding correction in relation to spillovers; the corrected estimate of long distance (6km) spillovers is no longer statistically significant, but the finding of within-school spillovers, upon which I rely (as well as the 3km spillovers, upon which I don't), are unchanged (Hamory Hicks, Kremer and Miguel 2015).

measures.

2.3 Critical periods

Despite reducing worm infection and improving school attendance, the original intervention appeared not to improve either measures of general intelligence in the long run or academic test scores in the short run for the *direct* recipients of the deworming medication. Part of the reason for this may be that for some types of outcomes, the deworming intervention came too late for participants in the original study: they were already of school age. The crucial phases for some aspects of both physical and cognitive development are thought to be within the first two or three years of life (Grantham-McGregor, *et al.* 2007, Knudsen, *et al.* 2006); nutrition shocks and changes to environmental stimuli in this period matter much more than they do later in life.[9] A few recent studies use rainfall changes to measure this effect. Hoddinott and Kinsey (2001) find that children in Zimbabwe who are malnourished between the ages of one and two because of a drought remain permanently 1.5-2 cm shorter than their counterparts who were not exposed to the same conditions; a follow-up paper shows that early-life reductions in height eventually manifest in lower height later in life, and lower educational attainment (Alderman, Hoddinott and Kinsey 2006a); older children exposed to the drought do not seem to suffer long-term harm. Maccini and Yang (2009) investigate long-term effects of good rainfall on children in Indonesia, and find that girls born in an area receiving 20 percent more annual rainfall than usual gain an additional 0.57cm in adult height, and complete an additional 0.22 grades of school, compared to children whose regions did not receive such beneficial rains.[10] Rainfall in other years had no significant long-term consequences.

[9] Windows during which such outside influences have especially strong effects are referred to as "sensitive" periods (Knudsen 2004); when the consequences are permanent, these periods are referred to as "critical." But because "critical" and "sensitive" periods differ across faculties (Knudsen, *et al.* 2006); I remain agnostic on whether de-worming could intervene in a particular "critical" period, relying instead on evidence that analogous early childhood interventions had substantial effects on health and education.

[10] Rainfall shocks at age two have similar (though statistically insignificant) effects on both outcomes. They do not find similar effects for men.

Panel studies have provided another empirical avenue for studying these effects: in Britain, cognitive skills at age seven predict around 20 percent of the variation in adult wages (Almond and Currie 2011a); in the US, parental income shocks in the first several years of a child's life matters much more for that child's eventual adult income than do shocks after the child's fifth birthday (Duncan, Ziol-Guest and Kalil 2010). In the US, variation in child height early in life predicts test scores later in life, even among children with the same mother (Case and Paxson 2010).[11]

Though unpredictable and extreme shocks periodically affect a small fraction of the population, less is known about whether policies—usually interventions aimed at addressing more mild conditions—can permanently alter human capital in this way. Still, a few exceptions stand out. Gertler, et al. (2014), provide a demonstration of the lasting socioeconomic benefits of early childhood interventions, by showing that an intensive early childhood stimulation intervention in Jamaica had large effects on eventual adult earnings. Similarly, in Guatemala, nutritional supplementation early in life led to earnings increases for men more than twenty years later (Hoddinott, Maluccio, Behrman, Flores and Martorell 2008). Barham (2012) finds large cognitive benefits to an early childhood health intervention in Bangladesh, measured when the children were between 8 and 14 years old. Improvements in early life health outcomes in the United States, brought about in part through racial integration of hospitals, are thought to have narrowed the black-white test score gap substantially once those healthier children became teenagers (Chay, Guryan and Mazumder 2009). Early childhood, however, is a particularly difficult time in a child's life from the perspective of policy: neither in the womb nor yet in school, this "sensitive" period falls beyond the reach of many government programs.[12]

Several very recent studies demonstrate links between early childhood de-

[11] Smith (2009) follows a similar approach, showing that in the US, variation in childhood health predicts subsequent household income and wealth, even within families.

[12] Field, Robles and Torero (2009) have shown that children who were *in utero* when their mothers received iodine supplementation eventually attain more schooling than siblings who did not benefit from the iodine. That timing is not the focus of the present paper, however.

worming and health, including four studies in East Africa all documenting short-term health gains. Alderman *et al.* (2006b), for example, found in a cluster-randomized trial that de-worming brings about weight improvements in pre-school-age children in Uganda, in a district that borders the PSDP study area around Lake Victoria. Children in the Uganda study were between 1 and 7 years old, but the study did not disaggregate effects by age; however, a study by Stoltzfus *et al.* (2004) in Zanzibar did. They show that children who were treated when less than 30 months old gained the most.[13] Within this young cohort, incidence of mild wasting[14] was cut nearly in half, from 36% in the control group to 18% in the treated group; older children did not improve nearly as much. The authors took note of this surprising aspect of their results: "The benefits thus occurred in the age group at highest risk for anemia and growth retardation, but in the age group with the lowest intensity of helminth infections."

The literature thus lays the groundwork for the present study. The simple question I ask is whether children who were infants or not yet born at the time of school-based deworming in Kenya received spillover benefits from the original PSDP intervention, by experiencing early childhood in a low-worm-infection environment. They were at the right age for long-term impacts to be large; recent literature suggests that worm infections are important even in the first year outside the womb; and the PSDP study showed that children who were simply near the dewormed schools also benefited from spillovers.

2.4 Long-term studies of early deworming

This paper is not alone in its long-term study of early childhood deworming. A parallel contribution in this vein is the working paper by Croke (2014), who conducts a long-term follow-up of the Alderman, et al. (2006) study by matching study areas to recently collected academic-subject-specific per-

[13] The Stoltzfus study was randomized at the level of "block" groupings four children.

[14] *Mild wasting:* having weight-for-height worse than one standard deviation below average, $WHZ < -1$

formance data in Uganda.[15] Croke demonstrates higher mathematics and English scores for children given deworming medication early in life; effect sizes are comparable to those here, though the mechanics of the intervention and the nature of the outcome variables differ somewhat. In that case, deworming medication was delivered as a component of "child health days" in Uganda; in the present case, deworming was delivered at school. Given the attendance rates involved, in any given round of deworming, this produced a lower compliance in Uganda than in the present study environment in Kenya. In Uganda, the target age of medication recipients was between 1 and 7 years old; in the present case, it was school-age children. This means that Croke estimates a long-run direct effect, while the present case relies on spillovers. Finally, Croke's outcomes are centered on tests of academic skills rather than underlying cognitive abilities. Though these are correlated capabilities, and though measures of the two are inextricably linked, the conclusions we draw might differ, either because of the nature of treatment, the context, or the measures. I discuss this further in the results section.

3 Data collection, 2009-2010

In 2009 and 2010, a field team in Kenya collected height, weight, migration, and basic family demographic data from more than 20,000 children at all of the deworming project schools in Samia and Bunyala districts of Kenya's Western Province.[16] For a subset of just over 2,400 children, the team also

[15]The dataset on academmic performance that Croke uses was collected independently by the Uwezo initiative.

[16]Here, I follow the original Primary School Deworming Program school lists; the team visited those that were not flooded (causing a temporary program disruption) at the time of the intervention. The original deworming program treatment assignments (Groups 1, 2, and 3) were not made salient to enumerators at any time during the data collection. The collection of data for this project was not a part of the original PSDP project, and was not in any way linked to the distribution or administration of deworming drugs. Enumerators were also blinded to the definition of treatment in this analysis, which involves the specific cohorts at specific schools; enumerator manipulation of outomes in relation to the definition of treatment is especially implausible because the enumerators measuring cognition and height were not the enumerators who recorded children's ages (See Appendix Section A.2).

conducted detailed cognitive assessments. Children from the same age cohorts were included during both data collection years: in 2009, this meant including every child between the ages of 8 and 14; in 2010, it meant every child between the ages of 9 and 15. These age cohorts were chosen both because they were still enrolled in primary school at the time of data collection, and because of how these cohorts align with the original intervention. Note that children were surveyed, not parents; this limited the range of questions on which data could be reliably gathered.

The randomized design of the original deworming project at the school level permits its use for estimation in this study, as shown in Figure 1: In the catchment areas of schools where deworming began in 1998, the children who were born in 1998 (and who I find in 2009 or 2010 as 11- or 12-year-olds) began experiencing the spillover effects of deworming in their year of birth; thus their age at the time of deworming, A_{id}, is less than one year (see Panel A, middle column, top row). I consider them "treated" for the purposes of the present study. Children from the same birth cohort but where deworming only began in 2001 had to wait until age three for school-based deworming to arrive; I consider them "untreated" for the purposes of the present study (see Panel A, middle column, bottom row). The figure outlines these definitions of treatment and comparison, by school and by treatment arm.[17] Because deworming started in different schools at different times, I can control for age at observation separately from age at the time of school treatment.

Summary statistics on the study population are shown in Table 1. Roughly half the sample is female, the average age is between 11 and 12, and average height is roughly what would be expected for these ages, if a bit low. Roughly 28 percent of the sample had migrated since birth. In-migration to these communities in response to Kenya's 2008 post-election violence[18] left school

[17]For the borderline case of children whose age was approximately 1 when deworming arrived, I consider them neither clearly treated nor untreated, as I only measure age up to a precision of one year increments, and the literature is not definitive on how these cases should be handled, had measurement been precise. Empirically, this is absorbed through an additional dummy variable, as discussed further below.

[18]A description of the post-election violence in relation to this geographic area can be found in Jakiela and Ozier (2015).

populations inflated with recent migrants from urban areas; for my results, I exclude those migrants from all regressions, since they were not present in these communities at the time of deworming in the late 1990s. Out-migration in response to the conflict is much less of a concern, since these rural areas are moderately ethnically homogeneous, and did not experience a high level of conflict.[19]

In Panels B and C of Table 1, I restrict attention to the sample of non-migrants. Panel B shows that the non-migrants are demographically much the same as the full sample, and goes on to tabulate several other characteristics: 21.6 percent of this population is stunted[20]; respondents had an average of 1.45 older siblings who attended the same primary school; 22.5 percent had at least three such siblings, while 37 percent had no older siblings who attended the same primary school.[21] These measures are used to assess the likely intensity of the deworming spillover effects, as discussed further in Section 5. Panel C simply shows the distribution of indicators for age at the time of the school deworming, explained in Figure 1.

In Panels D and E, I further restrict the sample to those for whom a cognitive survey was carried out. Because the cognitive survey takes roughly ten times as long as anthropometric measurement, the cognitive outcomes were gathered only for a random subsample of respondents.[22] Panel E shows that the characteristics of the respondents sampled for cognitive surveys do not substantially differ from the characteristics of all respondents.

The cognitive module included two measures of "verbal fluency," in which children name as many items in a category as they can in one minute. The first category is foods; the second is animals. The Peabody Picture Vocabulary Test (PPVT-4, Form B) measures "receptive vocabulary," in which children point to one of four pictures that best matches a word that has been read aloud to them. There are eighteen sets of questions in the test, each with twelve words; respondents proceed up through sets of increasing difficulty until they

[19]Nonetheless, I check for out-migration changes in Section 5.4.2.
[20]Stunting: height-for-age Z-score less than -2
[21]Questions on siblings were asked as a roster; key questions are in Appendix A.1.
[22]The procedure for randomly sampling respondents is described in Appendix A.2.

make nine mistakes in a single set, and are likely to be simply guessing. For reasoning, I use the 12-question Set B of J. C. Raven's Progressive Matrices, a series of puzzles commonly used to measure nonverbal reasoning and general intelligence.[23] For short-term memory, I use "digit-spans" of increasing length, in which respondents attempt to repeat a string of numbers back to the interviewer, either forwards or backwards. I provide raw means and standard deviations in Table 1, but for all regressions, I consider standardized versions of these cognitive measures, each re-scaled to have mean zero and standard deviation one in the study sample.[24]

Though it is not tabulated, I also condense these six measures using their first principal component in some parts of the analysis. Interpretation of coefficients on cognitive tests is clarified in Appendix Tables A2 through A5. The first column of Table A2 shows the weights on each outcome that yield the first principal component used in the analysis. Weights are almost equal across the different cognitive outcomes.[25] Because almost equal weight is given to each measure, I also construct a simple normalized sum of cognitive measures as an outcome to confirm the robustness of the findings where relevant. Correlations among cognitive measures are shown in Table A3: all are positive. To orient the reader, the cross-sectional relationships between cognitive performance, age, and grade in school are shown in Appendix Tables A4 and A5. In the

[23]Cattell (1971) and Raven (1989) describe the matrices and what they measure; of all the cognitive measures included in this study, performance on Raven's Matrices may be the most closely related to innate intelligence.

[24]Note that I standardize in the full sample rather than standardizing the measures based only on the distribution the comparison group, as one might do in a more conventional randomized trial. I do this because what I consider the "comparison group," depicted in Figure 1, is a group that is, on average, older than the "treatment group," though there are overlapping cohorts that provide identification. For the more difficult tests, this means that the younger respondents have mostly low responses, while the older respondents have a wider distribution of responses, so this aging alone changes the distributions slightly. Nonetheless, different scalings would not affect the statistical significance of the findings, nor would their magnitudes change substantially. For Raven's Matrices, for example, the standard deviation in the comparison group is 1.135 times that in the entire sample. For the normalized sum, it is 1.004 times that in the entire sample.

[25]The lowest weight is for "Verbal Fluency: Foods," perhaps the noisiest measure because it was the first exercise in the cognitive module. Low R^2 for regressions with this outcome also speak to its relative noisiness.

cross-section, coefficients on grade in school are typically one third larger than the coefficients on age, since pupils tend to repeat one grade out of every three. Conditional on grade in school, older children perform worse, since they are children who may have started school later or repeated grades more frequently.

4 Estimation

I begin discussion of estimation with a simple equation. For each individual i, consider that the relationship between an outcome, Y_i, and an indicator, $Before_i^C$, for whether that individual's school participated in mass deworming before the individual was C years old, is given by:

$$Y_i = \beta_1^C \cdot Before_i^C + \gamma_{2010} \cdot D_{Year_i=2010} + \epsilon_{1i} \tag{1}$$

Because of the original randomized design, conditional on age and data collection year, this type of exposure was actually, itself, randomized. So within a single birth cohort, controlling for the data collection year with a dummy variable, $D_{Year_i=2010}$, the equation above can be estimated direction. Second, aggregating across cohorts, and controlling for a set of interacted indicators for both year of data collection and respondent age:

$$Y_i = \beta_2^C \cdot Before_i^C + \sum_{A,Y} \gamma_{AY} D_{Age_i=A} \cdot D_{Year_i=Y} + \epsilon_{2i} \tag{2}$$

Estimating Equation 2 will provide more statistically powerful tests than would Equation 1, simply because it uses more of the available data; however, Equation 1 is a simple conceptual demonstration of the intended exercise. A challenge in estimating Equation 2 in the present environment is that children in this study report their age rounded to the year. From this type of data, I can roughly construct year of birth and thus, in relation to the randomized program rollout, the age at the time of school-based deworming. However, the equations above are specified in terms of whether deworming arrive in a child's local school before the child turned C years old. For the cohort whose survey

data suggest that their age was exactly C at the time of deworming, there is an equal probability that they really were or were not at least that exact age at the time deworming began. Direct estimation of β_2^C in equation 2 based on field data would thus be biased toward zero, because half of one cohort would be incorrectly categorized. This can be resolved, however, by including an indicator for reporting exactly a particular age at the time of deworming, and in general, though it diminishes statistical power to consider separately estimating the effect of spillovers from deworming arriving at each exact age for any single age or range of ages from C to C_H:

$$Y_i = \beta_3^C \cdot Before_i^C + \sum_{c=C}^{C_H} \beta_3^{ec} \cdot D_{A_{id}=c} + \sum_{A,Y} \gamma_{AY} D_{Age_i=a} \cdot D_{Year_i=Y} + \epsilon_{3i} \quad (3)$$

In case anything systematically differs for boys and girls in these communities and years, I can also absorb additional variation by separating the age and data collection year indicators by gender. Thus, for example, equation 3 above becomes:

$$Y_i = \beta_4^C \cdot Before_i^C + \beta_4^{eC} \cdot D_{A_{id}=C} + \sum_{A,S,Y} \gamma_{ASY} D_{Age_i=a} \cdot D_{Sex_i=S} \cdot D_{Year_i=Y} + \epsilon_{4i} \quad (4)$$

I do this for all the empirics that follow, though in practice, the results do not differ substantively if instead the gender interactions had been left out of the controls.

Because worm infections start only after a child is born, the earliest sensible value to consider for C is $C = 1$.[26] In what follows, I will consider a child "treated" with early deworming if her community's school started receiving deworming treatment before she was one year old.

[26]See Figure 3 from Brooker, Peshu, Warn, Mosobo, Guyatt, Marsh and Snow (1999), for example, for the pattern of infection using two datasets from Kenya and four from Asia, and Gyorkos, Maheu-Giroux, Casapía, Joseph and Creed-Kanashiro (2011) for data and discussion of analogous patterns in in Latin America. There is an additional potential mechanism regarding the role of hookworm in exacerbating anemia in pregnant women, referenced in Bethony, Brooker, Albonico, Geiger, Loukas, Diemert and Hotez (2006), though I do not explore that here.

In clarifying what this specification means, it is worth pointing out what it doesn't mean. There are certainly benefits to school-based mass deworming beyond age two, and all of the subjects in these datasets benefited in that way: by the time they were school-age, deworming was present in every school in this study. So to bring the analytical framework above to the variation in treatment timing in the study site in western Kenya is to arrive at a *lower bound* on deworming benefits: this approach estimates the differential benefit of particularly early deworming spillovers. Though it is the earliest sensible cutoff, $C = 1$ is not necessarily the "true" cutoff for a critical period in relation to deworming spillovers. The present data offer only limited variation to explore whether this cutoff appears sharp, or has more of a dose-response structure under some threshold age. I discuss alternative specifications in relation to this question (and thus assumptions about the relevant cutoff, and what those assumptions would yield) in Section 5.2 and in Table A1.

An alternative specification, based on Bleakley (2007), is to consider exposure to early childhood deworming in years:

$$Y_i = \beta_5^E \cdot Exp_i^E + \sum_{A,S,Y} \gamma_{ASY} D_{Age_i=a} \cdot D_{Sex_i=S} \cdot D_{Year_i=Y} + \epsilon_{5i} \qquad (5)$$

Above, exposure (Exp_i^E) is measured in years between birth and age E in which school-based deworming took place. I discuss this as a specification and robustness test in Section 5.2.

5 Results

Results in this study are clear enough that they can be measured without aggregating cohorts at all, following Equation 1. For birth cohorts in which school-based deworming arrived before age one for children in some communities, but afterwards for children in other communities, this estimation can be accomplished by simply conducting within-cohort regressions of outcomes on indicators for the original Miguel and Kremer (2004) study arms, including only a data collection year indicator as a control variable. I begin by docu-

menting effects on a measure of general non-verbal reasoning: correct answers to a series of visual puzzles from Raven's Progressive Matrices.

This is demonstrated in Figure 1. In Panel A, I explain the alignment of deworming timing to birth cohorts, to clarify which cohorts permit the relevant within-cohort comparisons. In Panel B1, I carry out those comparisons and show them using shaded bars; in Panel B2, I aggregate the effects to form a single coefficient estimate.[27] The effect of deworming spillovers early in life on nonverbal reasoning ten years later, measured in this very simple framework, aggregates to just under 0.3 standard deviations across all cohorts; this is larger than the impact of many education interventions.

The within-cohort estimation strategy makes transparent that, because of the original Miguel and Kremer (2004) randomized phase-in design, within-cohort estimation of effects in the present study treat the last arm of their study to start deworming ("Group 3") as the comparison group. If children in that group of schools were systematically different from those in the others, the results I show might simply be spurious. However, the estimation strategy lends itself to immediate falsification tests in relation to this hypothesis: checking whether there are systematic differences between the original study arms in cohorts where the absence of differential early deworming suggests that there should not be. As shown in Panels B1 and B2, differences are uniformly larger and more positive in the true measurement of effects than in the falsification tests. Comparing true tests of this paper's hypothesis with their falsification analogs, two out of three low-power within-cohort two-arm true tests are statistically significant in the estimation of Equation (1); zero of seven analogous falsification tests are significant.[28] These falsification tests serve two purposes: they provide additional certainty that the original randomization yielded balanced study arms; and whether or not it did, reinforces that the present study has an additional source of identification beyond the

[27] A full set of all pairwise treatment-arm-within-cohort comparisons is provided in Appendix Figure A1.

[28] I am also able to use previous years' academic achievement data from these schools to test whether there are systematic differences across treatment arms. I discuss these data and show the associated robustness checks in Section 5.4.4 and Tables A11 and A11.

original one. Here, treatment is defined by the interaction of the Miguel and Kremer study arms with birth cohorts.[29]

While this figure provides a direct window into the patterns in the data, results from estimating Equation 4 in a standard regression framework are shown in Tables 2 and 3. In Table 2, I report the estimated the β_4^1 coefficient, so that school-based deworming before (but not including) age 1 is compared with school-based deworming after (but not including) age 1.[30] Each row in the table reports $\hat{\beta}_4^1$ from a separate regression for a different outcome variable, Y_i. The effects are striking: school-based deworming in a child's community before a child's first birthday brings about a 0.2-standard-deviation improvement in performance in non-verbal reasoning (Raven's Progressive Matrices), a decade after the intervention, with a p-value less than 0.01 Estimated effects on vocabulary measures are similar in magnitude, but not always as significant; effects on memory are not statistically distinguishable from zero. A summary measure, the first principal component of all six cognitive measurements, also shows a roughly 0.2-standard-deviation effect, significant at the five percent level.[31]

One way to benchmark these effects is to compare them to the cross-sectional association between grade in school and cognitive measures; these relationships are shown in Appendix Tables A4 and A5. An additional grade in school is associated with an increase of roughly 0.4 standard deviations in the overall (first principal component) measure, and an increase of roughly 0.25 standard deviations in general reasoning (Raven's Matrices). Consider-

[29]Checking for balance in other ways is made complicated by the stepped design, since, for example, children are older, on average, in 2009 and 2010 in the cells that I consider "comparison." than in the cells that I consider "treated." However, I can do the equivalent of balance checks vis-a-vis the original Miguel and Kremer (2004) study arms (groups 1, 2, and 3) conditional on the data collection year: (A) whether the original study arm influences the gender balance; (B) whether the original study arm influences likelihood of being born in a particular year; (C) whether, conditional on being born in a particular year, the original study arm influences the likelihood of having at least three older siblings. Of these 63 statistical tests of balance, only three are significant at the 5 percent level.

[30]All estimation in this paper is conducted using Stata, versions 11.2, 12.1, or 13.1.

[31]An alternative formulation, the normalized sum of cognitive meeasures, shows a nearly identical coefficient magnitude and level of significance.

ing the impact of early childhood deworming spillovers on the first principal component and reasonsing measures, I take the ratio of coefficients in Table 2 to those in Appendix Tables A4 and A5. Thus, the effects of early deworming spillovers that I document are comparable to between 0.5 and 0.8 additional grades in school. That Raven's Matrices are so responsive to the intervention suggests that even mild disease burdens early in childhood can alter cognitive development. One of the key issues in the child development literature is the decreasing plasticity of physiological and neural development with age. The early age at which spillover effects of school-based deworming can impact child cognition has not been documented before, and may shed light on child development more generally.

While these cognitive effects are robust to a number of specifications, the effect of spillovers from school-based deworming on height, height-for-age, and stunting all appear statistically indistinguishable from zero. These estimates may be thought of as lower bounds, because even respondents in the excluded (comparison) group lived in communities that received treatment starting when they were aged two and older, and thus still may have experienced some beneficial effects.[32] As a robustness check (albeit with low power), in Table A10, I also check whether there are anthropometric effects among the subsample for whom cognitive measurements were carried out. Point estimates do not change very much, and again, none is statistically significant.

The absence of effects on stature is less surprising than it might appear at first glance, for several reasons. First, a literature suggests that in the face of inadequate food intake, a child's body "conserves energy by first limiting ... cognitive development, ... then by limiting the energy available for growth." (Lewit and Kerrebrock 1997, Center on Hunger, Poverty and Nutrition Policy 1995) A recent study of a vaccination program in the Philippines, for example, has found exactly this pattern (Bloom, Canning and Shenoy 2012). Second, a well-known effect of reducing worm infections is the concomitant reduction of

[32]Extreme stunting, defined as height-for-age Z score below -3, occurs with a frequency of roughly 4 percent in the sample. As with other measures of stature, there is also no discernible effect on extreme stunting.

anemia. However, while hemoglobin levels may affect cognitive development, they may not directly affect stature. Third, only 22 percent of the population in this study experiences stunting.[33] Other populations are not so lucky. For example, among children over 30 months old in a deworming study in Zanzibar, 41 percent were stunted - a rate almost twice as high as in the present study (Stoltzfus, et al., 2004); at the beginning of the INCAP study in Guatemala, 45 percent of 3-year-old children were measured as *severely* stunted, with a height-for-age Z-score below -3 (Hoddinott et al. 2008); in the present study, only 4 percent of the sample is this severely stunted - less than one-tenth the rate in Guatemala. In fact, in Jamaica, being stunted was an eligibility criterion for participation in the intervention studied by Gertler, et al. (2014). Certainly in relation to these well-known studies, malnutrition thus does not appear to be as severe a problem for this population, and as such, eliminating worm infections is less likely to have dramatic impacts on stature.

5.1 Heterogeneity and mechanisms

To help untangle the mechanisms behind this large effect on cognition, I consider different subpopulations in Table 3. I begin in the first column by repeating the specification shown in earlier tables, for reference. No matter what the mechanism, one might expect the spillovers to be larger within a household where older siblings receive treatment at school than in a household without such older siblings. Respondents were generally not certain of the ages of their older siblings, but as a simple rule, I consider those with at least three older

[33]Superficially, the Demographic and Health Surveys from Kenya appear to disagree with the stunting rate I measure: Table 11.1 of the 2008-2009 DHS report shows that 35.3 percent of Kenyan children are stunted; 34.2 percent in Western Province are stunted (Kenya National Bureau of Statistics 2010). However, the underlying data reveal that the rate of stunting in the two DHS clusters nearest to the site of the present study is actually 21.2 percent (author's calculations): almost exactly the same level I measure in this original data collection. Though the precision of averages based on only a few DHS clusters is low, one can marginally reject (at the 10 percent level) that the rate of stunting in these two clusters is equal to the national or provincial means of 34 or 35 percent; intra-cluster correlation in stunting rates is significantly different from zero, even after accounting for province fixed effects. Thus, calculations based on DHS data do (weakly) corroborate the stunting rates I observe.

siblings attending the same primary school to have had a sibling in school at the time of the deworming campaign.[34] When the sample is restricted to this group, shown in column 2, the point estimate of the effect size nearly doubles, though the difference is not statistically significant.

This raises the question of whether there are any spillovers for children who did not have siblings in the primary school that participated in deworming. If so, an epidemiological mechanism is supported; if not, a behavioral or financial within-household mechanism might be more plausible. Again, because of the imprecision of responses, I consider only respondents who did not have any older siblings attending the relevant primary school as the subsample best suited to answer this question; estimates are shown in column 3.[35] The effect is similar in magnitude to that of the full sample, and while for Raven's Matrices it is statistically significant, it is not for the first principal component of all cognitive measures. With this, evidence leans in favor of an epidemiological mechanism: fewer worms in schoolchildren means fewer worms in the community, and thus mean fewer infections in early childhood for these respondents.[36]

To further explore the sibling sample in column 2, I divide that group into those who had more female than male older siblings at the same primary school in column 4, and vice-versa in column 5. Sample size is quite small at this point, and standard errors widen, but it appears that the benefit is largest for those with older sisters at the primary school rather than older brothers.[37]

[34]Questions on siblings were asked as a roster, with key questions shown in Appendix A.1.

[35]Those who did not have any older siblings attending the same primary school may still have older siblings, but must report that none of them attended the same primary school. Thus, columns 2 and 3 are mutually exclusive but do not cover all categories; children who report that exactly one or two of their older siblings attended the same primary school are not included in either column.

[36]One can also test for effects at varying distances from Lake Victoria; results suggest effects both within 5km of the lake and beyond 5km from the lake, making geohelminths likely to be involved in the mechanism rather than exclusively schistosomes. This intuitively aligns with the more localized spillovers one would expect from soil-transmitted worms, as discussed in Miguel and Kremer (2004).

[37]Note that girls 13 years old and older in the original deworming study were excluded from receiving deworming medication because, at the time, it was not known whether deworming

The coefficient on Raven's Matrices in column 4 is significantly different from those in column 5 and column 1 at the five percent and one percent levels, respectively. The coefficients on the first principal component and normalized sum of cognitive scores in column 4 are significantly different from those in column 1 at the one percent and five percent levels, respectively.

This may reflect the relative frequencies with which girls and boys are tasked with caring for younger siblings: care for infants and toddlers by older female children is common in this study area, where the predominant ethnic group is Luhya. Weisner, et al. (1977) and subsequent authors have discussed how this care pattern is common across many cultures, and is salient in Kenya in particular. Weisner and co-authors call out the Luhya as a culture in which this pattern is especially strong: a high fraction of interactions among children are caretaking interactions with infants, and older female siblings were more than twice as likely to act as caretakers for infants as were their male counterparts (ibid., p. 175).

This pattern suggests that those who are in frequent physical contact with infants could be a key channel through which worms, or their absence through treatment, can affect infants and toddlers. Thus, this pattern provides further evidence in favor of an epidemiological mechanism. An alternative story could be that of a household budget constraint, in which healthier, dewormed older siblings would loosen budget constraints through reduced direct and indirect health costs, thereby freeing total resources to be devoted to the younger child. But in that story, health costs that determine the budget constraint would arise from both male and female older siblings. The pattern in the data seem to provide evidence against this story, as dewormed older male siblings appear not to have an impact on the younger ones in this sample.

medications had teratogenic effects (now, it is thought that they do not). All younger girls were given deworming treatment. Because birth spacing in this area is typically not very wide, older female siblings of children in the present sample would almost all have been dewormed. DHS data from Kenya, for example, show that for children with at least three older siblings, the sibling three birth orders older is typically 7 or 8 years old at the time of the birth of the child in question; for more than 90 percent of cases, this sibling is 12 years old or younger. As such, the three older siblings referred to in this study would have been dewormed in the vast majority of cases.

Finally, since a number of shocks and interventions in developing countries have been shown to have gender-specific impacts, I split the sample according to the sex of the respondent in columns 6 and 7. The coefficients are not appreciably (or statistically) different for boys and girls, though they are slightly higher for girls. This suggests that there is no substitution towards or away from any other nutrition or stimulation input that would be gender-specific, as is sometimes seen for interventions at later ages (Pitt, Rosenzweig and Hassan 2012).

5.1.1 Health channel versus a purely cognitive mechanism

Instead of acting through reduced worm infections and thus improved infant and toddler health, another way school-based deworming could have improved outcomes for the population in this study is through increased cognitive stimulation via more cognitively engaging play with smarter, healthier, older siblings, who received deworming at school directly. Because I did not observe the health of present respondents in their first years of life, or the level of stimulation they received from their siblings and neighbors, I cannot directly measure either channel. The cognitive channel described here would require increased stimulation (or, perhaps equivalently, quality of stimulation) from older children. Miguel and Kremer (2004) did not find any short-term test score improvements in the older children who were actually dewormed, so this channel is not highly plausible at face value. However, I can go further, and empirically test the plausibility of the cognitive channel by taking advantage of a feature of this study area: part of this area also benefitted from the textbook distribution program described in Glewwe, Kremer and Moulin (2009).

In the "School Assistance Program" (SAP) program that Glewwe et al. (2009) studied, the timing of resources delivered to schools varied by study arm. The key variants in their analysis were "SAP group 1," in which textbooks were distributed to schools in 1996, and the "comparison," or "SAP group 4," in which grants were delivered to schools in 2000. High-performing school-age children exhibited better academic performance in the short term as a consequence of the textbook distribution, as Glewwe et al. (2009) show

in their Table 8. Thus, if their younger siblings and neighbors benefited from a spillover of this increased cognitive stimulation, examining the cohorts born around these years should provide the test. Whether these younger cohorts did, indeed, benefit, is in turn, evidence for or against the plausibility of a purely cognitive channel (as opposed to a health channel) for the deworming spillovers that are the topic of this paper.

The SAP schools have some intersection with the deworming program schools; roughly one quarter of the sample surveyed for this paper attends a school that is included in either "SAP group 1" or "SAP group 4." Thus, in this dataset, I can exploit the fact that from 1996 through 1999, the "SAP group 1" schools were differentially advantaged in relation to the "SAP group 4" schools. To test the cognitive spillovers that would be most analogous to the present analysis of deworming spillovers, I examine cohorts which were no more than one year old by that time, following the specification below:

$$Y_i = \beta_6^{SAP} \cdot SAPcohort_i \cdot SAP1_i + \gamma_{G1} \cdot SAP1_i + \gamma_{cohort} \cdot SAPcohort_i + \sum_{A,S,Y} \gamma_{ASY} D_{Age_i=a} \cdot D_{Sex_i=S} \cdot D_{Year_i=Y} + \epsilon_{6i} \quad (6)$$

In equation 6, $SAP1_i$ is an indicator for whether the individual is in an "SAP group 1" school, as opposed to an "SAP group 4" school; the sample for this estimation is restricted to those two groups. The indicator $SAPcohort_i$ designates whether the individual is born in a cohort that would have received differential spillovers from SAP (more on this below). The coefficient of interest, β_6^{SAP}, is on the interaction of these two indicators. As before, fixed effects for the interaction of age, sex, and data collection year are also included.

Before proceeding to the empirics, I note that I face a decision about which cohorts to designate as having (plausibly) received differential spillovers from SAP, since the advantaged position of "SAP group 1" schools lasted from 1996 through 1999. To be analogous to the analysis of deworming spillovers by the first year of life, I can either consider those who were born either between 1995

and 1998 (no more than *just under one year old* when textbooks arrived, but also no more than that age when the grants arrived in the comparison schools), or between 1994 and 1997 (no more than *two years old*). I show results for both versions of this indicator.

In Table A9, for those children in schools that benefitted from SAP's textbook and grant distribution (specifically, groups 1 and 4), I test whether the arrival of textbooks by the first year of a child's life improves outcomes today. Though the subsample of schools leaves this test with only modest power, I find no consistent (or statistically significant) evidence of this pattern.

Thus, health remains the most plausible channel for deworming spillovers.

5.2 Variations on the empirical specification

In Table A1, I show a variety of specifications based on variations of Equation 3. In the first seven columns, I vary the value of C, the age before which deworming took place, from negative two to positive four.[38] For the indicators of subsequent deworming, I set C_H to four, and as C increases, the number of terms in the summation of later deworming indicators decreases.

Several regularities appear across the first seven columns of the table. First, the coefficients on deworming before age C cannot be statistically distinguished from one another for the first five columns (*before age -2* through *before age 2*), but after that, the coefficients lose significance and fall in magnitude. Either β_3^3 or β_3^4 (columns 6 and 7) can be statistically rejected as being equal to any of the coefficients from earlier columns. Second, the latest exact age indicator to be statistically significant is always age zero (in columns 1 through 3); conversely, in the first four columns of the table, the earliest coefficient to be statistically insignificant because of its lower magnitude is always that for deworming at exactly age one.

By including as many later deworming indicators as I do in columns 1 through 7, however, I sacrifice statistical power by reducing the size of the omitted group. Because of the two patterns described above, I repeat the

[38] Deworming "before age 0" means deworming before birth; "before age -1" means more than one year before birth; and so on.

specification from column 4 ($C = 1$ year) in columns 8 through 11, but decreasing C_H across the columns, until the specification in column 11 is simply that of Equation 2. The specification in column 10 yields the coefficients shown earlier in Table 2.

An alternative reading of Table A1 is that rather than narrowly favoring the main specification, it suggests one where children dewormed in their first year of life receive some fixed fraction of the treatment effect (perhaps half). This pattern would be consistent with either some misreporting of age, or with a gentle tapering of the most sensitive period for this effect; the data here unfortunately do not allow these possibilities to be empirically distinguished from one another.

Yet a different approach is that of Bleakley (2007), who considers years of deworming exposure. In his case, he interacts the program exposure with baseline intensity of hookworm infection, and considers up to the first 19 years of life for exposure. In this case, baseline rates of worm infection would probably not have the degree of variation seen across the much larger geographic area that Bleakley examines, and would be noisily measured in any event, so instead, a simple measure of exposure is used in Equation 5. This is the number of years of school-based deworming that took place between the participant's birth and age E, for which I consider $E = 1$, $E = 2$, and $E = 3$. If deworming spillovers after year E of life have important effects, the equation estimates only a lower bound. A fall in coefficient magnitude moving from $E = 1$ to $E = 3$ would be consistent with the first year mattering more than subsequent years. Indeed, in Appendix Table A6, this is exactly what I find.

In summary, then, the evidence here corroborates the results in the earlier sections: school-based deworming in a child's community prior to age 1 brings about a 0.2-standard-deviation improvement in performance on Raven's Matrices later in life; deworming at age 1 may have some positive effect, but smaller, though this could simply be due to noisy measurement of child age; deworming after age 1 cannot be statistically distinguished from deworming much later.

5.3 Discussion of results

Others have also found effects of deworming on cognition, though typically only in the short term. An observational study by Jukes, *et al.* (2002) investigated the relationship between cognitive function and helminth infections among Tanzanian schoolchildren, and found that after controlling for potential confounds, heavy schistosome infection was associated with lower performance on tests of short-term memory, reaction time, and information processing. A double-blind medical trial by Nokes, *et al.* (1992) found that the administration of albendazole led to immediate gains in memory skills in a population of Jamaican schoolchildren infected with whipworm and roundworm, and an experimental de-worming study with Tanzanian schoolchildren in the same region as the 2002 observational study also found cognitive gains in response to de-worming (Grigorenko, *et al.* 2006). Bleakley (2007) provides historical anecdotes that corroborate these patterns.

That I find effects mainly on reasoning–and to some extent, vocabulary–rather than memory may speak to the differences between slowed cognitive development and the more immediate cognitive impairments brought about by concurrent disease. Memory improves with age, but based on the results in this study, seems to depend less on health in early child development. Reasoning, however, shows a long-term response to improved health in early childhood. That stature is not affected suggests that worms do not cause severe caloric deprivation in early childhood in this population; the low intensity of worm infections at this age, and the low rate of stunting in relation to that in some other geographic contexts, is in accord with this possibility.

Though the evidence here is consistent with an effect that was largest for those under one year of age at the time of school-based deworming, the empirical variations explored in section 5.2 do not completely rule out a more gentle decline in effect with age, nor do they rule out any benefit received by all study participants for having any deworming in childhood. Croke (2014), for example, finds long-term benefits on mathematics and English scores for children who were roughly between 1 and 7 years old at the time of direct deworming. This could either be an effect that is above and beyond the one that

I measure here, or both of our studies could be estimating closely related effects, with noisily-measured age complicating both of our analyses. For Croke, the intensity of treatment tapers off for both the youngest children and the oldest children in the study, so it is difficult in that setting to separate a relatively more sensitive period for intervention from the intensity of treatment that children receive. Bleakley (2007) analyzes a natural experiment, and finds long-term benefits of deworming in the United States that include improvements in school attendance and literacy, though he considers any treatment in the first two decades of individuals' lives. This study adds clear evidence that whatever the benefit of direct school-based deworming, the benefit of the less worm-infected community that results is felt most strongly by the youngest cohorts.

5.4 Threats to identification

5.4.1 Demographics

Changes in the composition of cohorts in this study that are due to variation in school-based deworming treatment by community could potentially confound the analysis. Changes of this sort could arise if deworming changed mortality rates, leaving disproportionately healthy children as survivors.[39] One could also imagine that if adults adjusted their fertility patterns in response to school-based mass deworming—in either direction—such adjustment might change the interpretation of estimated effects. Bleakley and Lange (2009), for example, document decreases in fertility in response to the Rockefeller Sanitary Commission deworming work in the US South. A simple approach to mortality and fertility is to test whether respondents exposed to spillovers from school-based deworming from birth have more or fewer siblings than those who were exposed only later. An analogous approach is to test whether the actual quantity of age-eligible respondents in each school systematically varies as a

[39]Note, however, that a mortality mechanism does not have empirical support from studies that have examined it directly; see Awasthi, Peto, Read, Richards, Pande, Bundy and DEVTA Team (2013).

function of deworming exposure. Tests of these hypotheses are shown in Table A7. I find no evidence of either pattern using either approach.

5.4.2 Migration

If school-based deworming induced out-migration differentially among the families of those treated earlier or later, the set of children I find in 2009 and 2010 might be differentially selected by treatment arm, possibly resulting in biased coefficient estimates. The ideal test would look for families who had lived in the study area at the time of the deworming intervention, and would check whether they had moved.

I construct a test for this by taking advantage of the first round of the Kenya Life Panel Survey (KLPS1), which took place from 2003 to 2005. Although it followed the older children who had originally been in school at the time of deworming, rather than these children's families, the youngest respondents in KLPS1 were still in primary school and (generally) living with their parents at the time of the follow-up, so their location tells us their parents' location. For example, those who were in their second year of primary school (Standard 2) in 1998 would have been in Standard 7, Standard 8, or the first Form of secondary school from 2003 to 2005, if they had never repeated a grade. Most students repeat at least one grade, however, meaning most of the youngest respondents were still in primary school.[40] The KLPS1 respondents were randomized into two waves of surveying, so I can also focus a test on the first wave (from 2003 to 2004) to be sure not to confuse a departure for secondary schooling or marriage for the migration of the respondent's parents.

In Table A8, I show four versions of this test, using either Standard 2 or both Standards 2 and 3, and using either the first wave of KLPS1 or both the first and second waves. The outcome variable is whether the KLPS1 interview took place outside Busia District. The mean is around six percent, but it does not vary by treatment arm in any specification, and the standard errors

[40]Of the pupils who had been in Standard 2 in 1998, and were interviewed as late as 2005, more than 95 percent were not yet in secondary school. Of 1998 Standard 2 pupils interviewed in 2003 and 2004, more than 99 percent were not yet in secondary school.

are relatively small (1.4 percentage points in Column 1, for example). In this way, I establish (with relatively good precision) that there was not differential out-migration of families by original treatment arm.

5.4.3 Attrition

One could also imagine differential health or academic performance inducing different attendance rates among those whose schools initially experienced deworming at different times. Such a pattern could bias coefficient estimates if different treatment groups attrit differentially from the sample via differential enrollment or differential absenteeism. Comparison of the enrollment rosters by original treatment arm does not reveal any statistically significant differences. Similarly, comparison of the number of respondents in each school, by original treatment arm, does not reveal any statistically significant differences. There is thus no indication that this concern is borne out by the data.

5.4.4 School academic characteristics

Although Figure 1 shows no evidence of systematic differences between the schools in the original three treatment arms in terms of outcomes of interest, it is possible that some part of the variation in cognitive abilities measured in this study is due to pre-existing differences in levels of academic ability that vary at the community level.

I first test whether there were differences, in this regard, using school-level average scores on the Standard 8 primary school leaving examination, the Kenya Certificate of Primary Education (KCPE). Tests of differences by PSDP treatment arm are shown in Table A11. The last column shows that, in the five years prior to the start of PSDP, there were no differences on average between the three arms. The first five columns show that none of fifteen pairwise tests of differences is significant at the 5 percent level, though Groups 2 and 3 differed slightly (significant at the 10 percent level) in a single year, 1994.

I then test whether these small variations are responsible for the patterns

attributed to deworming spillovers in this paper, by replicating the analysis shown in Table 2, but including five separate controls for KCPE scores in the five years before KCPE. Results are shown in Table A12. This robustness check involves discarding more than ten percent of observations, since I am missing some schools' KCPE records from the mid-1990s. Despite this reduction in sample size (and thus power), the magnitudes and patterns of statistical significance remain largely unchanged.

6 Cost-effectiveness

The policy implications of a program's impact depend on its cost-effectiveness. As is well-established, the costs of deworming a child directly are low: $0.59 per pupil per year (as discussed by Baird, et al., 2014). I examine only spillover impacts, so the effects documented here may be thought of as additional benefits beyond those documented by Baird, et al., without additional costs.

To quantify benefits, I extrapolate (as Baird, et al., do) earnings impacts of early changes in cognition. For this, I turn to a pair of developing country studies. Grantham-McGregor, et al. (1997) found that an early childhood stimulation intervention in Jamaica increased performance on a variety of measures by the time subjects were 8 years old. In a follow-up, Gertler, et al. (2014), find that this intervention eventually increased wages in young adulthood by 25 percent. Though Jamaica and Kenya are clearly dissimilar in many ways, the comparison should still provide an instructive example.

In Table 4, I compare coefficients on four cognitive measures in the Grantham-McGregor, et al. (1997) study to those in the present study. Multiplying their ratio by the 25 percent wage increase found by Gertler, et al. (2014) yields extrapolated wage increases. Ranging from 12.5 to 83.3 percent, even the smallest is quite substantial. I can then multiply by earnings to place this in dollar terms. Baird, et al., estimate the net present value (NPV) of typical lifetime earnings for an individual in western Kenya at $1509.96.[41]

[41] One could, instead, extrapolate lifetime earnings using Kenya's GNI per capita as reported by the 2014 World Development Indicators. At $860 in current (non-PPP) dollars

Because the effects shown here are on cognition rather than years of schooling, my calculation does not include any government-borne costs of additional schooling. The question is, by what fraction will lifetime earnings rise, and how much additional revenue will the government eventually collect as a consequence? Multiplying the NPV of earnings by the most conservative percentage from Table 4 yields an additional NPV of just over $180 in earnings. Multiplying by a 16.5 percent tax rate, I find an additional NPV of government revenue of just over $30 per child benefiting from spillovers.[42]

While this figure should be accurate for each child receiving spillovers, the last piece of this calculation is to consider how many children benefit from each child actually being dewormed. I consider a population in which all cohorts are of equal size. The Primary School Deworming Program dewormed eight cohorts of schoolchildren each year, while the spillover benefit was felt by the birth cohort that year. This reduces the benefit by a factor of eight, per year per pupil dewormed. For comparison to the cost-benefit calculus of the 1998-2001 program in Kenya, this figure should be scaled back up by the 2.41 years of deworming, on average, that each dewormed pupil received. This still yields between $2.25 and $9 of additional benefit per pupil dewormed in the original deworming program. These benefits increase the already substantial public finance benefits of roughly $13 per pupil dewormed (shown by Baird, et al.) by 17 or 70 percent, depending on the extrapolation method - all benefits reaped from a roughly $1 per person subsidy for deworming medication.

(World Bank 2014), the relevant steps would yield a larger total figure. Conservatively, I use the Baird, et al., approach.

[42] A simpler and potentially more conservative approach is to consider the effects in terms of years of schooling. As discussed above, the cognitive effects appear comparable to half a year of school. If the returns to education are, conservatively, six percentage points per year (Duflo 2001, Card 2001), and are due exclusively to the cognitive human capital that is accrued through schooling, then an appropriate calculation is the same as the one above, but with three percent rather than 12.5 percent of the NPV of lifetime earnings. The result is then a more modest $7.50 gain for public coffers.

7 Conclusion

In this study, I measure the effect of deworming spillovers during early childhood. I find improvements in cognitive performance equivalent to between 0.5 and 0.8 years of schooling. Effects are nearly twice as large for children with an older sibling likely to have received deworming medication directly. This bolsters theories of sensitive periods for cognitive development, and provides evidence that an inexpensive intervention can benefit children immensely at this time. In light of the patterns of heterogeneity seen in the data, the most plausible explanation appears to be an epidemiological spillover, transmitted to infants and toddlers via the dewormed schoolchildren who are their older siblings and neighbors.

In relation to deworming specifically, this evidence lends further support to expanding initiatives worldwide that treat deworming en masse. Taken together with the recent working papers by Baird, et al., (2014) and Croke (2014), this study helps paint a complete picture of long-run benefits of deworming in developing countries. This expanding body of evidence has already led to real policy initiatives. In Kenya, for example, national deworming was undertaken in 2009, and began recurring annually in 2012. Infection levels in Kenya have dropped substantially since the original Miguel and Kremer study began in 1998 (Mwandawiro, et al., 2013). However, high-intensity infections remain prevalent around Sub-Saharan Africa and the world.

More broadly, the evidence that early childhood health shocks have ramifications for subsequent human capital in a variety of forms builds on, and gives empirical substance to the models of Grossman (1972), Cunha and Heckman (2008), and their successors. The present study, demonstrating the presence of early childhood deworming spillovers, in essence asks whether this inexpensive health intervention can act as an input to early cognitive skills. As in these and other models of human capital formation, the present findings do not rule out the value of interventions later in life. In fact, a successful early childhood intervention such as this one might well be complementary to the more frequently-studied interventions that become relevant later in the life

cycle. As elements of policy, however, the cost-effectiveness of deworming in highly worm-infected settings, and potentially that of other early-life health and nutritional interventions, will be difficult to match.

8 References

Ahuja, Amrita, Sarah Baird, Joan Hamory Hicks, Michael Kremer, Edward Miguel, and Shawn Powers, "When should governments subsidize health? The case of mass deworming," Working Paper 21148, National Bureau of Economic Research May 2015.

Aiken, Alexander M, Calum Davey, James R Hargreaves, and Richard J Hayes, "Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a pure replication," *International Journal of Epidemiology*, July 2015.

Albonico, Marco, Henrietta Allen, Lester Chitsulo, Dirk Engels, Albis-Francesco Gabrielli, and Lorenzo Savioli, "Controlling Soil-Transmitted Helminthiasis in Pre-School-Age Children through Preventive Chemotherapy," *Public Library of Science: Neglected Tropical Diseases*, March 2008, *2* (3), e126.

Alderman, Harold, John Hoddinott, and Bill Kinsey, "Long term consequences of early childhood malnutrition," *Oxford Economic Papers*, 2006, *58* (3), 450–474.

___, Joseph Konde-Lule, Isaac Sebuliba, Donald Bundy, and Andrew Hall, "Effect on weight gain of routinely giving albendazole to preschool children during child health days in Uganda: cluster randomised controlled trial," *British Medical Journal*, June 21 2006, *333*, 122–127.

Almond, Douglas and Janet Currie, "Human Capital Development before Age Five," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4, Amsterdam: Elsevier Science Publishers B.V., 2011, Chapter 15, pp. 1315–1486.

___ and ___, "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, Summer 2011, *25* (3), 153–172.

Awasthi, Shelly, Richard Peto, Simon Read, Susan M Richards, Vinod Pande, Donald Bundy, and DEVTA Team, "Population deworming every 6 months with albendazole in 1 million pre-school children in north India: DEVTA, a cluster-randomised trial," *Lancet*, 2013, *381* (5), 1478–1486.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel, "Worms at Work: Public Finance Implications of a Child Health Investment," *University of California at Berkeley*, 2014, - *mimeo*.

Barham, Tania, "Enhancing Cognitive Functioning: Medium-Term Effects of a Health and Family Planning Program in Matlab," *American Economic Journal: Applied Economics*, 2012, *4* (1), 245–273.

Bethony, Jeffrey, Simon Brooker, Marco Albonico, Stefan M. Geiger, Alex Loukas, David Diemert, and Peter J. Hotez, "Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm," *Lancet*, May 6 2006, *367* (9521), 1521–1532.

Bleakley, Hoyt, "Disease and development: evidence from hookworm eradication in the American south," *The Quarterly Journal of Economics*, February 2007, *122* (1), 73–117.

___ **and Fabian Lange**, "Chronic disease burden and the interaction of education, fertility, and growth," *The Review of Economics and Statistics*, February 2009, *91* (1), 52–65.

Bloom, David E, David Canning, and Erica S Shenoy, "The effect of vaccination on children's physical and cognitive development in the Philippines," *Applied Economics*, 2012, *44* (21), 2777–2783.

Brooker, S., E. A. Miguel, S. Moulin, A. I. Luoba, D. A. P. Bundy, and M. Kremer, "Epidemiology of single and multiple species of helminth infections among school children in Busia District, Kenya," *East African Medical Journal*, 2000, *77* (3), 157–161.

___ , **N. Peshu, P. A. Warn, M. Mosobo, H. L. Guyatt, K. Marsh, and R. W. Snow**, "The epidemiology of hookworm infection and its contribution to anaemia among pre-school children on the Kenyan coast," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, May-June 1999, *93* (3), 240–246.

Bundy, D. A. P., "Population Ecology of Intestinal Helminth Infections in Human Communities," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 1988, *321* (1207), 405–420.

___ **et al.**, "Intestinal nematode infections," in Christopher J. L. Murray, Alan D. Lopez, and Colin D. Mathers, eds., *The Global Epidemiology of Infectious Diseases*, Cambridge: Harvard University Press, 2001, Chapter 9, pp. 243–300.

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

___ , **M. S. Wong, L. L. Lewis, and J. Horton**, "Control of geohelminths by delivery of targeted chemotherapy through schools," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 1990, *84*, 115–120.

Bundy, Donald A. P., Judd L. Walson, and Kristie L. Watkins, "Worms, wisdom, and wealth: why deworming can make economic sense," *Trends in Parasitology*, March 2013, *29* (3), 142–148.

Card, David, "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 2001, *69* (5), 1127–1160.

Case, Ann and Christina Paxson, "Causes and Consequences of Early-Life Health," *Demography*, 2010, *47* (S), S65–S85.

Cattell, Raymond B., *Abilities: Their Structure, Growth, and Action*, Boston: Houghton Mifflin Company, 1971.

Center on Hunger, Poverty and Nutrition Policy, "The Link Between Nutrition and Cognitive Development in Children," *Tufts University School of Nutrition, Medford, MA*, 1995.

Chay, Kenneth Y., Jonathan Guryan, and Bhashkar Mazumder, "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth," Working Paper 15078, National Bureau of Economic Research June 2009.

Croke, Kevin, "The long run effects of early childhood deworming on literacy and numeracy: Evidence from Uganda," *Harvard School of Public Health*, 2014, - *mimeo.*

Cunha, Flavio and James J. Heckman, "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 2008, *43* (4), 738–782.

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 2001, *91* (4), 795–813.

Duncan, Greg J., Kathleen M. Ziol-Guest, and Ariel Kalil, "Early-Childhood Poverty and Adult Attainment, Behavior, and Health," *Child Development*, 2010, *81* (1), 306–325.

Dupas, Pascaline, "Health Behavior in Developing Countries," *Annual Review of Economics*, 2011, *3* (1), 425–449.

Evans, David, "Worm Wars: The Anthology," *Development Impact Blog*, August 2015.

___ **and Anna Popova**, "What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews," *World Bank Policy Research Working Paper*, 2015, (7203).

Field, Erica, Omar Robles, and Maximo Torero, "Iodine Deficiency and Schooling Attainment in Tanzania," *American Economic Journal: Applied Economics*, 2009, *1* (4), 140–169.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor, "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 2014, *344* (6187), 998–1001.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin, "Many Children Left Behind? Textbooks and Test Scores in Kenya.," *American Economic Journal: Applied Economics*, 2009, *1* (1), 112–135.

Grantham-McGregor, Sally M., Susan P. Walker, Susan M. Chang, and Christine A. Powell, "Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children.," *The American Journal of Clinical Nutrition*, 1997, *66* (2), 247–253.

Grantham-McGregor, Sally, Yin Bun Cheung, Santiago Cueto, Paul Glewwe, Linda Richter, Barbara Strupp, and the International Child Development Steering Group, "Child development in developing countries 1: Developmental potential in the first 5 years for children in developing countries," *Lancet*, January 6 2007, *369*, 60–70.

Grigorenko, Elena L., Robert J. Sternberg, Matthew Jukes, Katie Alcock, Jane Lambo, Damaris Ngorosho, Catherine Nokes, and Donald A. Bundy, "Effects of antiparasitic treatment on dynamically and statically tested cognitive skills over time," *Journal of Applied Developmental Psychology*, 2006, *27* (6), 499–526.

Grossman, Michael, "On the Concept of Health Capital and the Demands for Health," *Journal of Political Economy*, 1972, *80* (2), 223–255.

Gyorkos, Theresa W., Mathieu Maheu-Giroux, Martín Casapía, Serene A. Joseph, and Hilary Creed-Kanashiro, "Stunting and helminth infection in early preschool-age children in a resource-poor community in the Amazon lowlands of Peru," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, April 2011, *105* (4), 204–208.

Hamory Hicks, Joan, Michael Kremer, and Edward Miguel, "Commentary: Deworming externalities and schooling impacts in Kenya: a comment on Aiken etal. (2015) and Davey etal. (2015)," *International Journal of Epidemiology*, July 2015.

Hoddinott, John and Bill Kinsey, "Child growth in the time of drought," *Oxford Bulletin of Economics and Statistics*, 2001, *63* (4), 409–436.

___, John A Maluccio, Jere R Behrman, Rafael Flores, and Reynaldo Martorell, "Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults," *The Lancet*, 2008, *371*, 411–416.

Hotez, Peter J., Donald A. P. Bundy, Kathleen Beegle, Simon Brooker, Lesley Drake, Nilanthi de Silva, Antonio Montresor, Dirk Engels, Matthew Jukes, Lester Chitsulo, Jeffrey Chow, Ramanan Laxminarayan, Catherine M. Michaud, Jeff Bethony, Rodrigo Correa-Oliveira, Xiao Shu-Hua, Alan Fenwick, and Lorenzo Savioli, "Helminth Infections: Soil-Transmitted Helminth Infections and Schistosomiasis," in Dean T. Jamison, Joel G. Breman, Anthony R. Measham, George Alleyne, Mariam Claeson, David B. Evans, Prabhat Jha, Anne Mills, and Philip Musgrove, eds., *Disease Control Priorities in Developing Countries*, 2nd ed., New York: Oxford University Press, 2006, Chapter 24, pp. 467–482.

Jakiela, Pamela and Owen Ozier, "The impact of violence on individual risk preferences: evidence from a natural experiment," *World Bank Policy Research Working Paper*, 2015, (7440).

Jukes, M. C. H., C. A. Nokes, K. J. Alcock, J. K. Lambo, C. Kihamia, N. Ngorosho, A. Mbise, W. Lorri, E. Yona, L. Mwanri, A. D. Baddeley, A. Hall, and D. A. P. Bundy, "Heavy schistosomiasis associated with poor short-term memory and slower reaction times in Tanzanian schoolchildren," *Tropical Medicine and International Health*, February 2002, *7* (2), 104–117.

Kenya National Bureau of Statistics, *Kenya Demographic and Health Survey 2008-09*, Nairobi: Kenya National Bureau of Statistics, 2010.

Knudsen, Eric I., "Sensitive Periods in the Development of the Brain and Behavior," *Journal of Cognitive Neuroscience*, 2004, *16* (8), 1412–1425.

____, James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff, "Economic, neurobiological, and behavioral perspectives on building America's future workforce," *Proceedings of the National Academy of Sciences*, July 5 2006, *103* (27), 10155–10162.

Lewit, Eugene M and Nancy Kerrebrock, "Population-based growth stunting," *The Future of Children*, 1997, pp. 149–156.

Maccini, Sharon and Dean Yang, "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *American Economic Review*, 2009, *99* (3), 1006–1026.

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

Mott, K. E., "Schistosomiasis," in Christopher J. L. Murray, Alan D. Lopez, and Colin D. Mathers, eds., *The Global Epidemiology of Infectious Diseases*, Cambridge: Harvard University Press, 2001, Chapter 12, pp. 349–391.

Mwandawiro, Charles S, Birgit Nikolay, Jimmy H Kihara, Owen Ozier, Dunstan A Mukoko, Mariam T Mwanje, Anna Hakobyan,

Rachel L Pullan, Simon J Brooker, and Sammy M Njenga, "Monitoring and evaluating the impact of national school-based deworming in Kenya: study design and baseline results," *Parasites and Vectors*, 2013, *6* (1), 198.

Nokes, C., S. M. Grantham-McGregor, A. W. Sawyer, E. S. Cooper, and D. A. P. Bundy, "Parasitic Helminth Infection and Cognitive Function in School Children," *Proceedings of the Royal Society of London, Series B - Biological Sciences*, February 1992, *247*, 77–81.

Ozier, Owen, "Perils of simulation: parallel streams and the case of stata's rnormal command," *World Bank Policy Research Working Paper*, 2012, (6278).

Pitt, Mark M., Mark R. Rosenzweig, and Mohammad Nazmul Hassan, "Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy," *American Economic Review*, 2012, *107* (7), 3531–3560.

Raven, John, "The Raven Progressive Matrices: A Review of National Norming Studies and Ethnic and Socioeconomic Variation Within the United States," *Journal of Educational Measurement*, 1989, *26* (1), 1–16.

Smith, James P., "The Impact of Childhood Health on Adult Labor Market Outcomes," *Review of Economics and Statistics*, 2009, *91* (3), 478–489.

Stoltzfus, Rebecca J. et al., "Low Dose Daily Iron Supplementation Improves Iron Status and Appetite but Not Anemia, whereas Quarterly Anthelminthic Treatment Improves Growth, Appetite and Anemia in Zanzibari Preschool Children," *Journal of Nutrition*, 2004, *134* (2), 348–356.

Taylor-Robinson, David C., Nicola Maayan, Karla Soares-Weiser, Sarah Donegan, and Paul Garner, "Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin and school performance," *Cochrane Database of Systematic Reviews*, 2015, (7), Art. No.: CD000371.

Weisner, Thomas S et al., "My brother's keeper: Child and sibling caretaking [and comments and reply]," *Current Anthropology*, 1977, pp. 169–190.

World Bank, *The Little Data Book*, International Bank for Reconstruction and Development, 2014.

Figure 1: Identification strategy, and key comparisons in raw data

Panel A defines "treatment" for this study. Each column represents a birth cohort; each row, a group of communities where school-based deworming began in a specific year. Shading corresponds to treatment status, defined in terms of child age when deworming began, A_{id}: dark gray indicates younger than one ("treated"); white indicates older than one ("untreated"); and light gray indicates one year old.

Panels B1 and B2 show the within-cohort differences between "treated" and "untreated" groups in terms of scores on Raven's Matrices questions, standardized within the sample.

Panel B1 shows comparisons for cohorts aligned with those in Panel A. Dark gray bars indicate the treatment effect, the within-cohort difference between "treated" and "untreated" groups, as defined above. White bars indicate "placebo tests," in which two groups in a given cohort have the same treatment status as defined in Panel A. An "X" symbol appears wherever the difference would involve a group categorized as neither "treated" nor "untreated." A full set of pairwise comparisons is shown in the appendix; all comparisons are combined in a regression framework and are discussed further in the text.

Panel B2 shows the result of aggregating these pairwise comparisons in a simple regression: the dark bars yield to a coefficient of 0.294 (p=0.002), while aggregating the light bars yield a coefficient of 0.015 (p=0.82), as shown.

In Panels B1 and B2, the only control is a data collection year dummy (interacted with cohort dummies in B2), and the sample is restricted to non-migrants. 95 percent confidence intervals are shown, clustered by school-cohort.

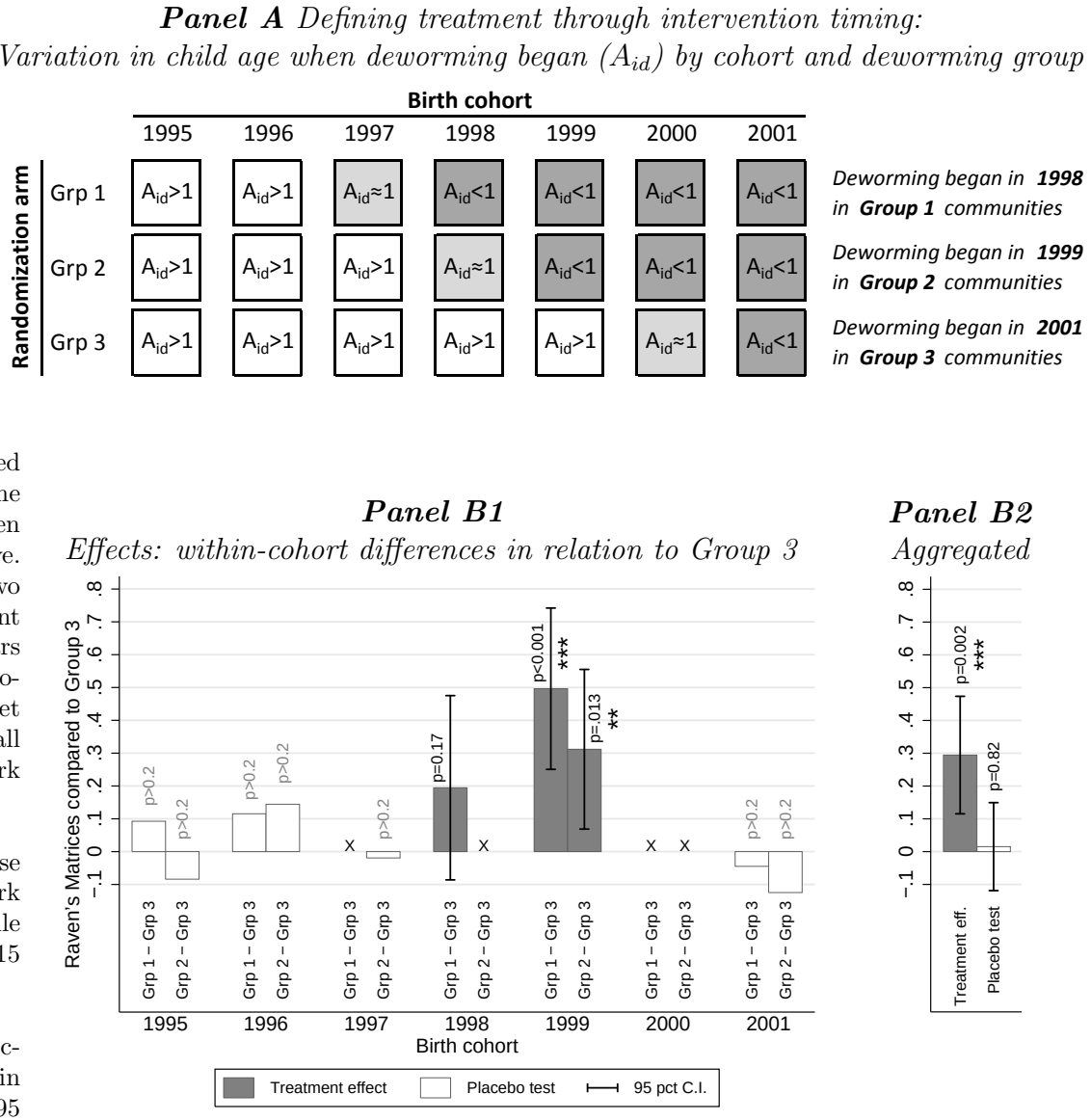


Table 1: Summary Statistics

Panel A: Characteristics, unconditional			
Characteristic	Mean	Standard Dev.	N
Age	11.488	(1.952)	21309
Female	0.497	(0.500)	21309
Height (cm)	141.560	(12.658)	21309
Ever migrated	0.281	(0.449)	21309
Panel B: Characteristics, conditional on non-migration and complete data			
Age	11.386	(1.953)	15158
Female	0.481	(0.500)	15158
Height (cm)	140.889	(12.694)	15158
Stunting (WHO 2007 HAZ< -2)	0.220	(0.414)	15158
Older siblings at same school	1.451	(1.593)	15158
At least 3 such siblings	0.224	(0.417)	15158
No such siblings	0.370	(0.483)	15158
Panel C: Deworming cohort, conditional on non-migration and complete data			
Deworming before age -1	0.165	(0.371)	15158
Deworming starting at age -1	0.118	(0.322)	15158
Deworming starting at age 0	0.129	(0.335)	15158
Deworming starting at age 1	0.145	(0.353)	15158
Deworming starting at age 2	0.151	(0.358)	15158
Deworming starting after age 2	0.292	(0.455)	15158
Panel D: Cognitive data, conditional on non-migration and complete data			
Verbal Fluency: Foods	9.243	(2.950)	2424
Verbal Fluency: Animals	8.856	(3.222)	2424
Vocabulary: highest PPVT level	6.037	(3.328)	2421
Reasoning: Raven's Matrices	3.624	(1.932)	2423
Memory: Digit Span Forwards	3.352	(1.746)	2406
Memory: Digit Span Backwards	0.952	(1.241)	2371
*Panel E: Characteristics, conditional on non-migration and **cognitive** data*			
Age	11.530	(1.927)	2371
Female	0.471	(0.499)	2371
Height (cm)	141.728	(12.888)	2371
Stunting (WHO 2007 HAZ< -2)	0.213	(0.410)	2371
Older siblings at same school	1.456	(1.631)	2371
At least 3 such siblings	0.225	(0.418)	2371
No such siblings	0.377	(0.485)	2371

Table 2: Main effects: spillovers from school-based deworming before age one

Outcome	**Effect**
Raven's Matrices	0.211***
	(0.079)
PPVT Level	0.169*
	(0.097)
Verbal fluency	0.200**
	(0.091)
Memory: digit span forwards	0.128
	(0.096)
Memory: digit span backwards	0.022
	(0.089)
All cognitive: First principal component	0.215**
	(0.099)
All cognitive: Normalized sum	0.215**
	(0.098)
Height (cm)	0.210
	(0.298)
Height-for-age z-score	0.030
	(0.044)
Stunting (HAZ<-2)	0.001
	(0.015)

In the table above, the excluded group comprises the cohorts whose communities experienced school-based deworming during their second year of life or later. Each coefficient comes from a separate regression of the indicated outcome on indicators for the age at deworming. Standard errors are clustered at the school-cohort level; gender×age×data collection year fixed effects are included. All cognitive outcomes are standardized (variance=1). Only non-migrants are included in this analysis. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3: Spillover effects of school-based deworming before age one: different subpopulations

Subpopulation: / Outcome:	[1] Full sample	[2] With older siblings[a]	[3] Without older siblings[a]	[4] Female siblings[b]	[5] Male siblings[b]	[6] Female[c]	[7] Male[c]
Raven's Matrices	0.211***	0.440***	0.238**	0.842***	0.103	0.220*	0.198
	(0.079)	(0.162)	(0.117)	(0.268)	(0.195)	(0.114)	(0.126)
All cognitive: First PC	0.215**	0.408**	0.186	0.771***	0.261	0.249**	0.180
	(0.099)	(0.158)	(0.134)	(0.254)	(0.237)	(0.121)	(0.136)
All cognitive: Normalized sum	0.215**	0.395**	0.190	0.752***	0.272	0.246**	0.182
	(0.098)	(0.158)	(0.136)	(0.255)	(0.235)	(0.122)	(0.137)
Observations	2365	533	894	235	225	1113	1252

In the table above, the excluded group comprises the cohorts whose communities received school-based deworming during their second year of life or later. Each coefficient comes from a separate regression of the indicated outcome on indicators for the age at deworming. Standard errors are clustered at the school-cohort level; gender×age×data collection year fixed effects are included. All cognitive outcomes are standardized (variance=1). Only non-migrants are included in this analysis. Column [1] repeats the specification shown in Table 2, for reference. (a) In column [2], the sample is restricted to respondents who have at least three older siblings who attended the same primary school; in column [3], it is restricted to those for whom no older siblings attended the same primary school. (b) In column [4], the restriction is similar to that in column [2], but with the added restriction that more female than male older siblings attended the same primary school; in column [5], it is reversed: more male than female older siblings attended the same primary school. (c) In columns [6] and [7], the original sample is simply split according to the gender of the respondent. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 4: Extrapolating benefits of early cognition interventions

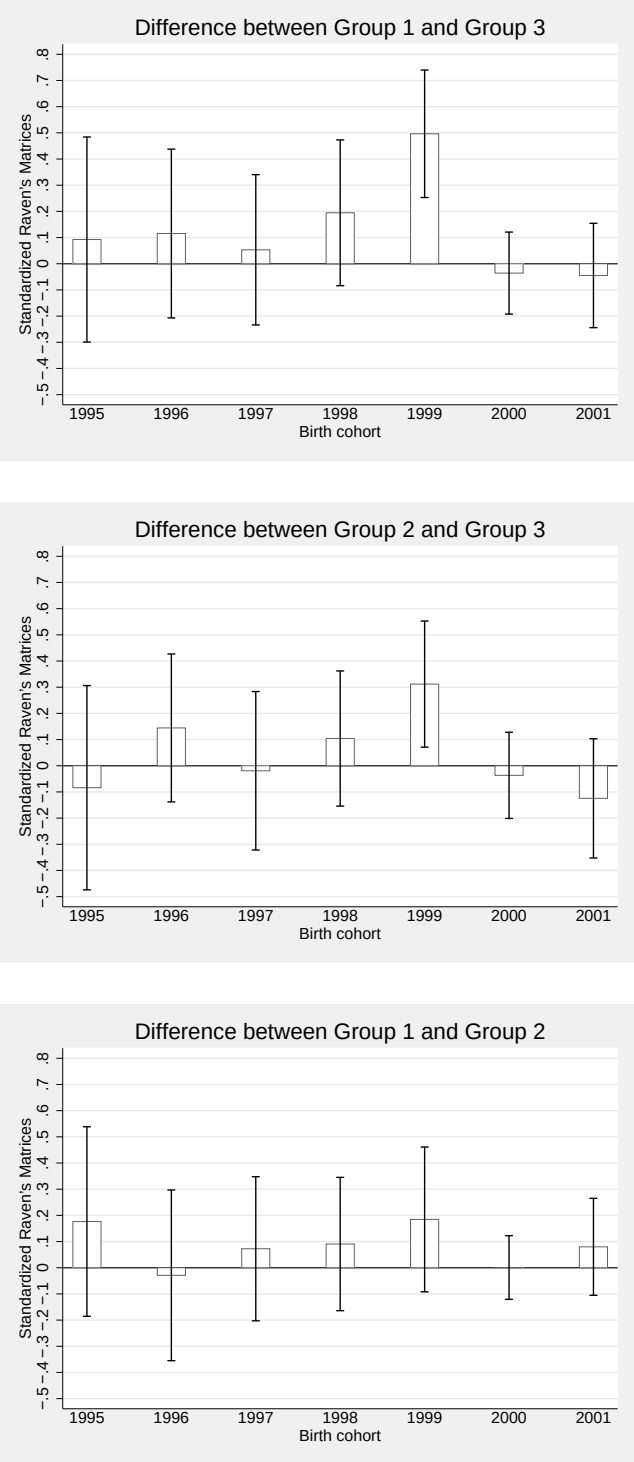
Cognitive measure	**Coefficients**		**Ratio**	**Wage change**
	1997 study (point est.)	**This study** (point est.)		(extrapolated pct.)
Raven's Matrices	0.86 questions	0.43 questions	0.497	12.5
Forward digit span	0.20 digits	0.12 digits	0.59	14.8
PPVT vocabulary	3.0 words	3.4 words	1.13	28.3
Verbal Fluency	0.3 answers	1.0 answers	3.33	83.3

The 1997 study is Grantham-McGregor, Walker, Chang and Powell (1997). Note that the measure of verbal fluency used by Grantham-McGregor, et al., included three categories of answers, each for one minute, while the present study only included two. As such, an alternative calculation would scale the coefficient in the present study by 1.5 for better comparability, yielding a coefficient ratio of 5, and a predicted wage increase of 125 percent.

A Appendix

FOR ONLINE PUBLICATION

Figure A1: Full set of comparisons between treatment arms



Note: Bars represent 95-percent confidence intervals.

Table A1: Locating the critical period: different simple specifications

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
Deworming before age -2	0.284*	.	.	.	.	.	.	.	.	.	.
	(0.148)										
Deworming before age -1	.	0.245*	.	.	.	.	.	.	.	.	.
		(0.134)									
Deworming before age 0	.	.	0.291**	.	.	.	.	.	.	.	.
			(0.132)								
Deworming before age 1	.	.	.	0.28**	.	.	.	0.273***	0.22**	0.22***	0.137**
				(0.127)				(0.105)	(0.092)	(0.078)	(0.066)
Deworming before age 2	.	.	.	.	0.229*	.	.	.	.	.	.
					(0.121)						
Deworming before age 3	.	.	.	.	.	0.11	.	.	.	.	.
						(0.112)					
Deworming before age 4	.	.	.	.	.	.	0.081	.	.	.	.
							(0.098)				
Deworming age -2	0.235*	.	.	.	.	.	.	.	.	.	.
	(0.135)										
Deworming age -1	0.316**	0.319**	.	.	.	.	.	.	.	.	.
	(0.136)	(0.136)									
Deworming age 0	0.27**	0.267**	0.273**	.	.	.	.	.	.	.	.
	(0.132)	(0.132)	(0.131)								
Deworming age 1	0.193	0.194	0.196	0.194	.	.	.	0.187*	0.139	0.139*	.
	(0.125)	(0.125)	(0.125)	(0.125)				(0.098)	(0.086)	(0.077)	
Deworming age 2	0.051	0.051	0.05	0.049	0.046	.	.	0.042	-0.0005	.	.
	(0.118)	(0.118)	(0.118)	(0.118)	(0.118)			(0.097)	(0.088)		
Deworming age 3	0.09	0.09	0.091	0.092	0.09	0.058	.	0.085	.	.	.
	(0.104)	(0.104)	(0.104)	(0.104)	(0.104)	(0.103)		(0.089)			
Deworming age 4	0.012	0.012	0.013	0.012	0.021	-0.008	-0.017	.	.	.	.
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.129)	(0.127)				
Observations	2472	2472	2472	2472	2472	2472	2472	2472	2472	2472	2472
R^2	0.135	0.135	0.135	0.135	0.134	0.133	0.132	0.135	0.135	0.135	0.133

In the table above, each column represents a separate regression with standardized performance on Raven's Matrices as the outcome variable. In columns [1]-[7], the omitted category is respondents for whom school-based deworming took place when they were already five years old or older. Because this is a relatively small group, columns [8]-[11] show the same estimation as in column [4], but with different omitted categories: arrival of school-based deworming after ages four and older; three and older; two and older; and one and older, respectively. Gender×age×data collection year fixed effects are included in all specifications, all samples are restricted to non-migrants, and standard errors are clustered at the school-cohort level.

Table A2: Cognitive measures: Principal Components

Principal component:	(1)	(2)	(3)	(4)	(5)	(6)
Verbal Fluency: Foods	0.3612	-0.6743	0.0027	0.2230	0.5550	-0.2390
Verbal Fluency: Animals	0.4443	-0.4238	-0.0030	-0.0594	-0.5293	0.5825
Digit Span Forwards	0.3814	0.2288	0.6677	-0.5286	0.2687	0.0693
Digit Span Backwards	0.3875	0.3937	0.2948	0.7742	-0.0915	-0.0117
Vocabulary: PPVT	0.4762	0.0878	-0.2600	-0.2420	-0.4023	-0.6910
Raven's Matrices	0.3870	0.3882	-0.6322	-0.0965	0.4115	0.3481
Explained variance:	0.4665	0.6214	0.7464	0.8482	0.9344	1.0000

Table A3: Cognitive measure correlations

	Fluency: Foods	Fluency: Animals	Digit Span Forwards	Digit Span Backwards	Raven's Matrices	Vocab: PPVT
Foods	1.0000					
Animals	0.5007	1.0000				
Digit Span Forwards	0.2400	0.3389	1.0000			
Digit Span Backwards	0.2323	0.3183	0.3778	1.0000		
Raven's Matrices	0.2218	0.3014	0.2742	0.3477	1.0000	
PPVT	0.3490	0.5204	0.3989	0.3899	0.5083	1.0000

Table A4: Cognitive performance (first principal component, normalized) as a function of observables

	All			Boys			Girls		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grade	0.451***	0.382***	.	0.459***	0.407***	.	0.449***	0.355***	.
	(0.011)	(0.007)		(0.015)	(0.009)		(0.016)	(0.01)	
Age	-0.089***	.	0.261***	-0.069***	.	0.292***	-0.118***	.	0.226***
	(0.011)		(0.009)	(0.015)		(0.012)	(0.015)		(0.012)
Constant	-0.872***	-1.607***	-3.021***	-1.071***	-1.652***	-3.369***	-0.606***	-1.554***	-2.624***
	(0.095)	(0.032)	(0.103)	(0.134)	(0.043)	(0.146)	(0.133)	(0.048)	(0.145)
Observations	2583	2583	2585	1372	1372	1373	1203	1203	1204
R^2	0.555	0.543	0.254	0.582	0.576	0.287	0.532	0.51	0.218

Table A5: Cognitive performance (normalized) as a function of observables

	Outcome											
	Vocabulary: PPVT		Verbal fluency: Foods		Verbal fluency: Animals		Memory: Digit Span Forwards		Memory: Digit Span Backwards		Reasoning: Raven's Matrices	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Grade	0.372***	.	0.196***	.	0.279***	.	0.219***	.	0.222***	.	0.247***	.
	(0.007)		(0.009)		(0.008)		(0.009)		(0.009)		(0.009)	
Age	.	0.261***	.	0.143***	.	0.212***	.	0.118***	.	0.139***	.	0.17***
		(0.009)		(0.01)		(0.009)		(0.01)		(0.01)		(0.009)
Constant	-1.565***	-3.012***	-0.81***	-1.642***	-1.169***	-2.444***	-0.918***	-1.363***	-0.936***	-1.608***	-1.034***	-1.960***
	(0.032)	(0.101)	(0.042)	(0.112)	(0.039)	(0.107)	(0.042)	(0.115)	(0.043)	(0.115)	(0.041)	(0.111)
Observations	2661	2665	2664	2667	2664	2667	2633	2635	2591	2593	2663	2667
R^2	0.519	0.255	0.145	0.078	0.292	0.168	0.179	0.052	0.184	0.072	0.227	0.107

Table A6: Linear effects of years of deworming before different ages

Outcome:	Before age 1 [1]	Before age 2 [2]	Before age 3 [3]
Raven's Matrices	0.137**	0.110***	0.076**
	(0.066)	(0.039)	(0.031)
All cognitive: First PC	0.164**	0.105**	0.069*
	(0.077)	(0.048)	(0.037)
All cognitive: Normalized sum	0.164**	0.105**	0.069*
	(0.078)	(0.048)	(0.036)
Observations	2412	2412	2412

The table above presents an alternative specification. Here, each coefficient comes from a separate regression of the indicated outcome on years of school-based deworming in a child's community between that child's birth and a particular age (Equation 5). Thus in column 1, the independent variable takes the value 0 or 1; in column 2, it is either 0, 1, or 2; and in column 3, it ranges from 0 to 3. As usual, standard errors are clustered at the school-cohort level; gender×age×data collection year fixed effects are included; all cognitive outcomes are standardized (variance=1); only non-migrants are included.

Table A7: Testing for fertility or mortality responses to deworming

	Sibling N	Respondent N
	[1]	[2]
Deworming before 1 (main specification)	0.056	0.062
	(0.055)	(0.799)
Deworming before birth (alternative for fertility)	0.061	0.862
	(0.049)	(0.832)
Observations	15630	1740

The table above presents tests of a fertility response to school-based deworming in the community. In the first column, the outcome variable is the number of younger siblings reported by the respondent. In the second column, observations have been aggregated at the level of the { data collection year × birth year × gender × migration indicator × school }. Thus, in the second column, the outcome is simply the count of observations in these bins. The first row presents the same specification as elsewhere in the paper, showing an indicator for school-based deworming arriving in the community in the respondent's year of birth or earlier; the second row presents an alternative specification, using an indicator for whether deworming arrived before the respondent was born. In any of the four cells, a significant coefficient could indicate a change in fertility in response to mass school-based deworming starting in that community. In the first column, standard errors are clustered at the school-cohort level. In the second column, because observations are already aggregated, standard errors are simply heteroskedasticity-robust.

Table A8: Testing for parent migration response to deworming

	KLPS Waves 1 and 2		KLPS Wave 1 only	
1998 Class:	*Std2 or Std3*	*Std2*	*Std2 or Std3*	*Std2*
	[1]	[2]	[3]	[4]
Group 2	-0.006	-0.001	-0.02	-0.006
	(0.014)	(0.018)	(0.019)	(0.028)
Group 3	0.0004	-0.003	0.005	-0.005
	(0.014)	(0.018)	(0.02)	(0.027)
Constant	0.064***	0.056***	0.068***	0.065***
	(0.01)	(0.013)	(0.014)	(0.019)
Joint F p-value	0.887	0.985	0.392	0.974
Group 2 = Group 3	0.662	0.923	0.202	0.979
Observations	1871	924	924	454
R^2	0.0001	0.00003	0.002	0.0001

The table above presents tests of a migration response to school-based deworming. The outcome variable is constructed from the first round of the Kenya Life Panel Survey, 2003-2005, and is an indicator for whether the respondent lived outside Busia at the time of the interview. The first two columns aggregate the two rounds of KLPS surveying; the third and fourth column restrict attention to the first round. The odd-numbered columns look at those who were enrolled in either Standard 2 or Standard 3 in 1998; the even-numbered columns look only at those who were in Standard 2 in 1998. Schools were grouped for deworming timing; this table shows coefficients on the indicators for being in a Group 2 or Group 3 school in 1998; Group 1, which received deworming first, is the excluded group. The "Joint F p-value" row indicates the p-value from the test that the two coefficients shown are different from zero. The "Group 2 = Group 3" row indicates the p-value from the test that the coefficients on Group 2 and Group 3 are equal. Heteroskedasticity-robust standard errors are in parentheses.

Table A9: Testing for spillovers on younger sibings from textbook distribution

Outcome:	Raven's Matrices		Principal Component	
Age encoding:	1994-1997	1995-1998	1994-1997	1995-1998
	[1]	[2]	[3]	[4]
Textbooks arrive early	−0.105	−0.038	0.145	0.058
	(0.144)	(0.132)	(0.149)	(0.138)
Observations	663	663	663	663

The table above presents tests of cognitive responses to the intervention described by Glewwe, Moulin, and Kremer (2009). This table is discussed in Section 5.1.1, and shows the β_6^{SAP} coefficient from estimation of Equation 6. The outcome in columns 1 and 2 is standardized Raven's Matrices; in columns 3 and 4, it is the first principal component of all cognitive outcome measures. The even-numbered columns differ from the odd-numbered columns in the way a "treated" cohort is defined. The textbook program began in "SAP group 1" schools in 1996; the comparison schools, "SAP group 4," received grants in 2000. Cohorts just under 1 year old at the start of the textbook intervention were born in 1995; those just over one year old were born in 1994. Likewise, since the comparison group was treated with grants in 2000, those just under one year old were were born in 1999; those just over were born in 1998. Thus, the odd-numbered columns show the coefficients on the interaction between being in SAP group 1 and being born in the years 1994-1997, cohorts who may have benefited from intervention spillovers in SAP group 1 but not yet in SAP group 4; the even-numbered columns show analogous results but with the indicator for cohorts born in years 1995-1998. Standard errors, in parentheses, are clustered at the school-cohort level.

Table A10: Main anthropometric effects: sample with cognitive data

Outcome	**Effect**
Height (cm)	0.320
	(0.745)
Height-for-age z-score	0.049
	(0.108)
Stunting (HAZ<-2)	−0.003
	(0.036)

The table above is analogous to the bottom panel of Table 2, but restricting the analysis to the subsample for whom cognitive data were collected. Standard errors are clustered at the school-cohort level; gender×age×data collection year fixed effects are included. Only non-migrants are included in this analysis.

Table A11: Mean KCPE differences by deworming group

KCPE Year	1993	1994	1995	1996	1997	Average
	[1]	[2]	[3]	[4]	[5]	[6]
Group 2	0.02	0.012	0.017	0.022	-0.01	0.011
	(0.014)	(0.016)	(0.019)	(0.019)	(0.014)	(0.013)
Group 3	0.007	-0.025	0.016	0.011	-0.006	-0.0006
	(0.016)	(0.018)	(0.018)	(0.017)	(0.014)	(0.014)
Constant	0.466***	0.485***	0.464***	0.478***	0.479***	0.475***
	(0.01)	(0.01)	(0.012)	(0.012)	(0.009)	(0.009)
Joint F p-value	0.347	0.153	0.57	0.519	0.768	0.666
Group 2 = Group 3	0.409	0.056	0.935	0.571	0.833	0.436
Observations	71	71	70	73	72	67
R^2	0.025	0.063	0.016	0.02	0.007	0.012

The table above presents tests of differential KCPE averages across schools. The KCPE is the Kenya Certificate of Primary Education, the primary school leaving examination taken at the end of Standard 8 (8th grade). The data available to me are missing some years' KCPE averages at some schools, so the number of observations varies from column to column. The outcome variable in each column is the school-level mean KCPE score (as a percent out of 700) in one of the years prior to the deworming program, except in Column 6, where the outcome is the average of the five KCPE means used in the first five columns. Schools were grouped for deworming timing; this table shows coefficients on the indicators for being in a Group 2 or Group 3 school in 1998; Group 1, which received deworming first, is the excluded group; mean KCPE percentage for Group 1 is given in the "Constant" row. The "Joint F p-value" row indicates the p-value from the test that the two coefficients shown are different from zero. The "Group 2 = Group 3" row indicates the p-value from the test that the coefficients on Group 2 and Group 3 are equal. Heteroskedasticity-robust standard errors are in parentheses.

Table A12: Main effects: robustness to five years of KCPE controls

Outcome	**Effect**
Raven's Matrices	0.194**
	(0.083)
PPVT Level	0.135*
	(0.078)
Verbal fluency	0.191**
	(0.093)
Memory: digit span forwards	0.114
	(0.103)
Memory: digit span backwards	0.098
	(0.080)
All cognitive: First principal component	0.210**
	(0.095)
All cognitive: Normalized sum	0.210**
	(0.096)
Height (cm)	−0.064
	(0.321)
Height-for-age z-score	−0.012
	(0.047)
Stunting (HAZ<-2)	0.006
	(0.017)

In the table above, the excluded group comprises the cohorts whose communities received school-based deworming during their second year of life or later. Each coefficient comes from a separate regression of the indicated outcome on indicators for the age at deworming. Standard errors are clustered at the school-cohort level; gender×age×data collection year fixed effects are included. All cognitive outcomes are standardized (variance=1). Only non-migrants are included in this analysis. The only difference between this table and Table 2 is the inclusion of five separate years of school-level KCPE means as control variables. Because I have incomplete KCPE data, however, this means omitting a number of schools for which I do not have complete KCPE data. The number of observations used for the regression in the first row, for example, is 2,422 in Table 2, but is only 2,171 here.

Appendices on details of surveying

A.1 Precise wording of survey questions

The "child survey" was administered to children aged 8-14 in 2009, and aged 9-15 in 2010. It was administed at school, meaning that many household characteristics could not be directly observed.[43] Key questions:

1.	What is your date of birth? (or age)?	*Ulizaliwa mwaka gani? Mwezi/tarehe gani? (umri / miaka mingapi?)*
2.	Have you ever lived in a different place from the one you live in now?	*Umewahi kuishi mahali pengine mbali na mahali unapoishi sasa?*
3.	Do you have any siblings who are currently living and share the same mother and father as you?	*Una dada na ndugu / kaka wowote ambao umezaliwa nao kwa baba na mama mmoja walio hai?*
3.B.	Sex	*Jinsia*
3.C.	Older/younger/twin?	*Mkubwa/mdogo/pacha?*
3.D.	Still in school?	*Bado anasoma?*
3.E.	Did the sibling ever attend this primary school?	*Aliwahi hudhuria shule ya msingi hii?*

A.2 Sampling respondents for cognitive module

In both 2009 and 2010, a random sample of respondents was chosen for cognitive tests, since these modules took roughly ten times as long as anthropometrics. In both years, random uniform draws were made centrally using Stata 10.1 to sample ID numbers for cognitive tests. This preceded changes in Stata's random number generator in 2011 (Ozier 2012). The draws were made multiple times in each year in order to prevent predictability in the relationship between ID numbers and sampling for cognitive testing across schools. The procedure differed slightly in 2009 and 2010. In 2009, roughly one in twenty respondents was sampled for cognitive data. In 2010, roughly one in four respondents was sampled for cognitive data, conditional on being a nonmigrant. In both years, ID numbers were serially assigned to respondents at the time of completing the child survey; the key questions in that survey are described in Appendix A.1. Following that survey by a first field team member,

[43]Because of the brevity and low risk of this work, and because the cognitive testing closely resembled normal school activity, the protocol entailed giving teachers copies of a "parent information sheet" to send home with children before the team's visits to schools to inform parents of the team's planned activities, and to provide them with the opportunity to opt out. This protocol, and the details of the sheet, were approved by KEMRI and UC Berkeley CPHS.

children were measured for height and weight by a second group from the field team; at this second stage, if the ID tag given to the child by the first team matched a list of those numbers randomly sampled for cognitive testing that day, the second group referred the child to a third enumerator for completion of the cognitive tests. Thus because the enrollment team was obliged to enroll participants and assign them ID numbers serially, and did not know, as they gave ID numbers to children, either which random ID number sampling list was used that day or, within it, what numbers would be chosen for cognitive testing (because this list of randomly selected numbers was not in that team's possession), the procedure achieved what some call "allocation concealment" vis-à-vis this sampling process.

A.3 Details on number of respondents per "arm"

Evans and Popova (2015) have pointed out that the inclusion of a few additional details in research papers, perhaps particularly those with complex designs, can ease not only the interpretation of the paper, but also the task of determining whether and how to include those papers in meta-analyses. With that in mind, below, I report a few key numbers with respect to sample sizes.

Data collection year:	2009			2010		
Schools visited:	37			36		
PSDP Group	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Schools visited:	9	9	19	16	14	6
N Interviewed:	3092	1965	4594	4600	5003	2055
N Non-migrant:	2394	1393	3256	3117	3612	1386
N Non-migrant, cognitive:	97	62	138	826	872	376

Cell sizes for main results, Table 2				
Outcome	$A_{id} < 1$ ("treated")	$A_{id} \approx 1$	$A_{id} > 1$ ("comparison")	Total
Raven's Matrices	1171	367	884	2422
PPVT Level	1172	366	882	2420
Verbal fluency	1172	367	884	2423
Memory: digit span forwards	1169	363	873	2405
Memory: digit span backwards	1150	356	864	2370
All cognitive: First principal component	1149	356	860	2365
All cognitive: Normalized sum	1149	356	860	2365
Height (cm)	6230	2204	6719	15153
Height-for-age z-score	6230	2204	6719	15153
Stunting (HAZ<-2)	6230	2204	6719	15153

Table A13: Duplicate of main effects (Table 2) for reference

Outcome	**Effect**
Raven's Matrices	0.211***
	(0.079)
PPVT Level	0.169*
	(0.097)
Verbal fluency	0.200**
	(0.091)
Memory: digit span forwards	0.128
	(0.096)
Memory: digit span backwards	0.022
	(0.089)
All cognitive: PC1	0.215**
	(0.099)
All cognitive: Normalized sum	0.215**
	(0.098)
Height (cm)	0.210
	(0.298)
Height-for-age z-score	0.030
	(0.044)
Stunting (HAZ<-2)	0.001
	(0.015)

Table A14: Main effects, but restricting sample to 1998 and 1999 birth cohorts only

Outcome	**Effect**
Raven's Matrices	0.279***
	(0.094)
PPVT Level	0.190
	(0.123)
Verbal fluency	0.232**
	(0.108)
Memory: digit span forwards	0.100
	(0.120)
Memory: digit span backwards	0.098
	(0.102)
All cognitive: PC1	0.255**
	(0.123)
All cognitive: Normalized sum	0.257**
	(0.122)
Height (cm)	0.342
	(0.346)
Height-for-age z-score	0.051
	(0.051)
Stunting (HAZ<-2)	−0.010
	(0.018)

Table A15: Subpopulations analysis: variations on Table 3

Panel A: Duplicate of Table 3 for reference

Subpopulation: / Outcome:	[1] Full sample	[2] With older siblings[a]	[3] Without older siblings[a]	[4] Female siblings[b]	[5] Male siblings[b]	[6] Female[c]	[7] Male[c]
Raven's Matrices	0.211***	0.440***	0.238**	0.842***	0.103	0.220*	0.198
	(0.079)	(0.162)	(0.117)	(0.268)	(0.195)	(0.114)	(0.126)
All cognitive: First PC	0.215**	0.408**	0.186	0.771***	0.261	0.249**	0.180
	(0.099)	(0.158)	(0.134)	(0.254)	(0.237)	(0.121)	(0.136)
All cognitive: Normalized sum	0.215**	0.395**	0.190	0.752***	0.272	0.246**	0.182
	(0.098)	(0.158)	(0.136)	(0.255)	(0.235)	(0.122)	(0.137)
Observations	2365	533	894	235	225	1113	1252

Panel B: Restricting sample to 1998 and 1999 birth cohorts only

Subpopulation: / Outcome:	Full sample	With older siblings[a]	Without older siblings[a]	Female siblings[b]	Male siblings[b]	Female[c]	Male[c]
Raven's Matrices	0.279***	0.766***	0.305**	1.341***	0.055	0.345**	0.217
	(0.094)	(0.186)	(0.132)	(0.231)	(0.210)	(0.136)	(0.152)
All cognitive: First PC	0.255**	0.567***	0.210	0.924***	0.254	0.277*	0.231
	(0.123)	(0.186)	(0.152)	(0.283)	(0.291)	(0.144)	(0.171)
All cognitive: Normalized sum	0.257**	0.577***	0.213	0.941***	0.280	0.274*	0.238
	(0.122)	(0.182)	(0.153)	(0.280)	(0.286)	(0.144)	(0.171)
Observations	685	138	268	65	54	311	374

Tables A14 and A15 show that the analysis is robust to restricting the sample to only the 1998 and 1999 birth cohorts, where within-cohort experimental variation includes both individuals who were not yet one year old, and who were more than one year old, when school-based deworming began in their communities.