Worm Eradication: Evaluating a Staggered National Program *

Michael Beeli

May 15, 2021

1 Abstract

Renewed attention has been brought to the challenge of containing and treating infectious diseases. As many as 1.5 billion individuals globally continue to suffer from Soil-Transmitted Helminths, highly infectious and transmissible parasites that can be easily treated with a single pill. Despite the ease of treatment for STHs however, many experts argue that take-up of the treatment remains extremely low, worsened by imperfect information, collective action problems, and the lack of universal programs. I explore this argument with a quasi-experimental evaluation of the first two rounds of the largest single-day public health program in the world, India's National Deworming day. Utilizing a difference-in-differences approach, I exploit the targeting of treatment towards public school students in the country with high levels of private school enrollment. I evaluate the differential impacts of National Deworming Day between Indian states with relatively higher levels of state capacity and public health preparedness, and comment on how these differences influenced the success of the treatment. I find that the first two rounds of National Deworming Day have a significant impact on the educational outcomes for a nationally representative sample of rural Indian children aged 11-16, and that the Indian government's determination of states' preparedness had a significant impact on the success of the treatment.

*Thanks to Preeti Manchanda and the ASER Centre for providing me with and acquainting me with data for this study. Special thank you to Robert Cumby for serving as an excellent advisor and mentor for this project. A big thank you to William Jack and Martin Ravallion for serving as a mentor throughout my education and research.

2 Introduction

Since the launch of Kenya's first deworming campaign in 1998, the deworming treatment as a development and public health intervention has expanded globally throughout Africa and Asia. India held its first National Deworming Day (NDD) in government schools and Anganwadis (rural child care centers) in cooperation with Evidence Action on February 10 2015 in 12 of its 36 state and union territories. The next year, February 2016, National Deworming Day was expanded to include government schools all 36 states and union territories, as well as a small number of private schools.

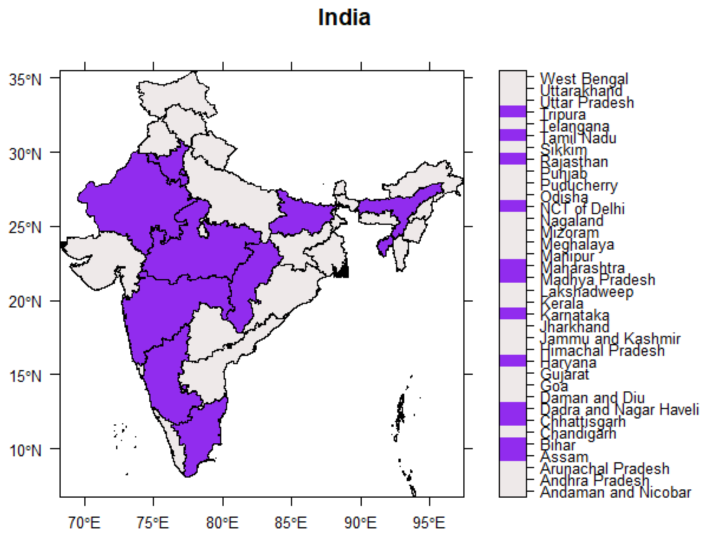


Figure 1: NDD: States Treated in 2015 — From Database of Global Administrative Areas

Miguel and Kremer (2004) showed that a school-wide deworming effort in Kenya had significant impacts on school attendance and public health of children in the treated schools, and due to the highly infectious nature of STHs, that the majority of the gains could be attributed to positive externalities. Ozier (2018) conducted a longitudinal study of the deworming externality and found impacts equivalent to 0.5 to 0.8 years of schooling for young children aged 1 at the time of the intervention in the region, and effects of double the size for younger siblings of treated children, further evidence of a large externality. Bleakley (2007) found a number of positive effects, including increased returns to schooling and long-run gains in income from a much older (1910) deworming campaign in the American South. Another study from Baird, Hicks, Kremer, and Miguel (2019) found the deworming treatment increased the pass rate for the Kenyan national primary school exit exam by female children by 9.5%.

While the majority of the literature on the subject in international development has been studied in the context of randomized experiments at the micro-level, this study aims to contribute to the existing literature measuring the impacts of a large-scale deworming treatment by specifically analyzing short-term impacts in India using a quasi-experimental design at the national level. Using national survey data on educational attainment (Math, English, and Reading scores) across private and public schools in the years 2014 and 2016, the study will exploit the exclusion of private schools to estimate the short-term impacts of India's first two national deworming days on public school children.

The study design requires consideration of public school children as a treatment group, and privately schooled children as a control group, with a minor compliance problem given the addition of a small number of private schools in February 2016's deworming day. As noted above, 12 of India's 36 states and union territories took part in NDD 2015 where only public schools were targeted for the treatment, and all 36 took part in 2016, where a small number of private schools were included. This comparison is possible due to high levels of private school enrollment in India.[1] Considering 2 levels of differences between school children — treated in 2015 or not, and public or private school attendees — a differences in differences estimation can be made under standard assumptions of parallel trends, or equivalent changes over time. Due to limited compliance of the control group (private school children may have received treatment at public schools, or attend the small number of treated private schools) and spillover effects of the treatment, the coefficients will likely be biased towards zero. The study will investigate whether significant coefficients can be found, despite the expected negative bias.

Holding the parallel trends assumption (and a few others that will be discussed later) true, a differences-in-difference design will allow a comparison of quasi-experimental treatment effects for public school children. Under such assumptions, a number of issues of endogeneity or selection bias in treatment status are effectively controlled for — guidelines for India's National Deworming Day indicate that the 12 states chosen for 2015's National Deworming Day were picked for their superior administrative capacity; however the nature of the design would not be affected by this fact if the parallel trends assumption holds. A similar line of reasoning follows for private and public school children, and the study will perform a number of analyses to verify the parallel trends assumption between the groups. Following this aggregate national analysis, I utilize a triple-differences approach to obtain results with increased precision and estimate differential impacts amongst the 12 states selected for their superior administrative capacity.

The merit of this paper is its novel approach to measuring the impacts of the deworming treatment. The nature of the randomized experiments used to evaluate the impacts of the treatment in the past several decades inherently limits the scope of the study to a relatively small micro-level. India's National Deworming Day is a valuable object of study given the intervention's sheer scale as well as its nature as a government-led campaign. NDD 2015 is notable for providing the treatment to over 90 million children in a single day, and NDD 2016 was reported to have reached over 170 million. This paper will, implicitly, attempt to answer the question of whether the Indian government's effort, the largest single-day public health campaign in the world, was a particularly successful program. Additionally, it will test if high-functioning states were more successful at carrying out the program. In doing so, the paper will

[1] Private school enrollment among Indian children was estimated at 52% in 2017 by the Central Square Foundation

add further contribution to the literature on the relative efficacy of universal government public health intervention, an increasingly critical unit of analysis in the study of development economics.

Data comes from ASER (Annual Status of Education Report) India's National Survey, which estimates a representative national sample of rural Indian children, including enrollment in government/private schools, math/reading/writing scores for children of different genders and ages across different states. The reports from 2014 and 2016 in particular will be used for estimation.

3 Background

3.1 Soil-Transmitted Helminths

Soil-Transmitted Helminths are known as not only one of the world's most common infections, but also one of the most harmful to public health, nutrition, educational attainment, and productivity. The World Health Organization estimates that as many as 1.5 billion, just under 25 percent of the global population, suffer from infection by one of many STHs (Pullan et al 2014). STHs fall under the more general category of Neglected Tropical Diseases, as many as 13 diseases and health afflictions that primarily affect the world's poorest individuals and remain largely untreated despite the existence of effective and somewhat affordable treatments in much of the world. The prevalence of these diseases is concentrated in areas afflicted by poverty, and with sub-optimal health care systems and delivery.

Infection has been repeatedly shown to have major adverse impacts on health and nutrition, including reduced calorie absorption, stunting, anemia and reduced birth weights for children of infected mothers. (Miguel & Kremer et al. 2019; Stephenson et al. 1993; Stoltzfus et al. 1997; Larocque at al 2006). Additional research has also shown a number of adverse immunological effects, including evidence that STH infection may reduce immunological capacity against other infections such as malaria. (Kirwan et al. 2010; Wammes et al 2016) A wide range of economic literature has contributed to knowledge of the negative impacts of STHs. Studies by Edward Miguel and Michael Kremer, Hoyt Bleakley, and Owen Ozier, to name a few, have shown positive effects of treatment for STHs on educational, nutritional, health, and labor market outcomes.

The emergence of the COVID-19 pandemic has brought renewed attention to the role of public health responses and infrastructure in a world where many preventable public health challenges remain. The prevalence of soil-transmitted helminths in particular, stands out as a particularly urgent public health, policy, and development issue.

3.2 Literature Review

Economic literature on the subject is anchored by proven effects on schooling, labor outcomes, health and nutrition. The deworming treatment, via mass distribution of Albendazole and Mebendazole, is of particular interest due to the low cost of treatment relative to potential gains. Even more important to STHs as a development issue however, is repeated evidence of significant externalities associated with treatment, implying insufficient demand relative to the socially optimal outcome. Miguel and Kremer et al. (2004) evaluated the effects of a randomized phased-in deworming program in Kenyan schools in a short-term study, finding that the program significantly reduced infections not only in treated schools, but also the surrounding area. Their seminal study showed that treated schools faced significantly lower rates of general infection among their student population, and communities with higher density of children living near treated schools saw major reductions in population adjusted infection rates.

Specifically, Miguel & Kremer et al. (2004) showed that the average within-school externality of treatment alone reduced moderate to heavy infection rates in treatment schools by 29 percentage points, and 9 percentage points in in non-treated schools. Accompanying the straightforward public health gains however, Miguel & Kremer also provide estimates for the deworming treatment's impact on educational outcomes. They show that school absenteeism was reduced by 25 percent, or 7.5 percentage points on average in treatment schools, and after accounting for externalities, estimate that the local average treatment effect on treated pupils is at least .14 additional years of schooling.

Given the impairing nature of STH infections, it is conceivable to find a wide variety of positive impacts associated with the deworming treatment. Because the program studied by Miguel & Kremer was targeted towards school-age children, it is useful to think of impacts on schooling as impacts on cognitive ability and general productivity, which in line with standard human capital theory (Mincer 1958), would lead to higher wages and better labor market outcomes.

Ozier (2018) followed in a similar manner, and took a deeper look at externalities associated with the same deworming program in Kenya. Ozier conducted a longitudinal study of a younger cohort of respondents from the same program from as Miguel and Kremer et al. (2004) who were ineligible to be directly treated, but were hypothesized to have benefited from positive externalities from the treatment. Ozier is able to identify children in treated areas who did not receive treatment directly, but are likely to have benefited from spillovers of the deworming treatment through reduced risk of infection. Given the young age of the children at the time of treatment, he hypothesizes that his sample's susceptibility to childhood impairment due to infection is relatively higher, and shows that younger siblings of treated children gained as many as 1-2 years of equivalent schooling due to the treatment.

Ozier's approach zeroes in further on the public health and childhood development dimension of the deworming treatment; specifically, he claims that his results corroborate theories of critical, sensitive periods in childhood development. Furthermore, his findings lend further support to the notion that low cost treatments early in life can yield significant high returns. These findings are also corroborated by Bleakley's (2007) study of a deworming campaign in the American South, which showed significant gains to children, but failed to show statistically significant gains for adults. Ozier's findings bear resemblance to a number of classic poverty reduction models in development economics, most notably the classic S-shaped poverty trap first pioneered by Mazumdar (1959) and Dasgupta and Ray (1986).

As parasitic worms, STHs greatly damage nutritional absorption, which could easily lead to nonlinearities in food consumption and realized nutrition as discussed by Ravallion (2013) among infected individuals that could even exacerbate an S-shaped nutritional model of poverty. STHs are known to feed on host tissue and blood, reducing iron and protein levels, increase malabsorption of key nutrients, and even cause a loss of appetite, diarrhea, or dysentery. Following Ravallion (2013) and Sachs (2005), large exogenous income gains may be needed to overcome such nutritional-based poverty traps, absent methods to remove or at least reduce the nutritional threshold of the trap. The low cost of treatment for most deworming infections then offers an invaluable tool in fighting nutritional poverty traps, yet researchers and policymakers have long wondered why private demand fails to meet socially optimal supply.

On one dimension, findings from Ozier (2018), Bleakley (2007), and Miguel & Kremer (2004, 2019) provide a straightforward answer vis-à-vis externalities associated with treatment. Yet, it is likely that the private gains to an individual greatly exceeds the cost of treatment as well. Bleakley's (2007) study found that deworming treatments provide not only increased schooling, but also a significantly higher return to schooling, as well as long-term income gains. Miguel & Kremer's (2019) 20-year follow-up of their original study published in 2004 found further evidence of an extremely private high rate of return to individuals who received 2-3 the deworming treatment, captured by a 14% increase in consumption expenditure, a 13% increase in hourly earnings, a 9% increase in non-agricultural work hours, and a 9% higher probability of living in urban areas. They provide a conservative social annual rate of return estimate of 37%.

Deworming treatments, then, provide a clear picture of a good with extreme levels of underconsumption, both at the social and private level. Ozier (2018) and Dupas (2011) discuss the missing demand for treatment with regards to externalities, and postulate that short-run discomfort or unfamiliarity associated with deworming can also play a key role in shaping preferences that may appear irrational to the policymaker. This research suggests that imperfect information regarding the effects and availability of the treatment, as well as an inflated discount factor, become challenges of interest. Other important market failures, such as transaction costs associated with delivery of treatment, as discussed by Singh (2008) and collective action problems barring public good provision as discussed in Weimann et al (2019) are also likely to play a key role in determining the efficacy of deworming treatments.

The primary mechanism of mitigating such market failures comes from government provision of public goods, which, following Acemoglu et al (2014) can provide the efficient Pigouvian subsidies, counteract imperfect private information, reduce transaction costs and solve collective action problems in public good delivery challenges. Acemoglu et al (2014) and Khemani (2019) consider the effects of state capacity on economic development, and characterize universal and efficient public good distribution as a key function of high state capacity. Acemouglu et al (2014)'s findings that State Capacity has significant impacts on public good delivery demonstrates the importance of state capacity in influencing the collective impact of National

Deworming Day.

Each robust evaluation of deworming programs has been able to successfully identify exogenous treatment of children (Miguel & Kremer et al. 2019). The 3 seminal papers on the subject all exploited the nature of Kenya's randomized phased-in deworming program in Southern Busia, from 1998 to 1999. Due to the randomized nature of the study's phased-in approach, Miguel & Kremer (2004) were able to estimate mean treatment effects with a simple estimation strategy, including time differences in treatment, with more sophisticated methods of decomposing both the direct effect and externality. Their 2019 follow-up uses a similar approach, utilizing an 84% tracking rate of individuals from the original study. Critically however, Miguel and Kremer use an Intention-to-Treat (ITT) framework in their estimation using their individual-level data, in order to capture the wide range of likely spillover effects. While simple externalities in terms of treatment are likely, Miguel & Kremer also bring attention to issues involving transfer bias (if children attempt to switch enrollment to gain access to treatment) or complementarities between treatment and school attendance. Specifically, if the increased school attendance of childrens' peers, due to their improvements in health status, plays a role in attracting more children to attend school, the impact of the deworming intervention could be understated by a standard Treatment on the Treated (ToT) approach.

Ozier (2018) is able to exploit exogenous variation in a similiar manner, Ozier evaluates the same study in Busia, while Bleakley (2007) uses a quasi-experimental strategy to account for the lack of a randomized experiment in its accompanying data for his study of deworming in the 20th-century American South. Bleakley exploits differences in incidence rates of hookworm across regions with differences in soil and climate, which allowed a treatment-control strategy due to significant differences in potential gains from the mass deworming campaign that took place. Bleakley constructs a reduced-form difference-in-differences estimator that exogenously identifies the impact of the eradication campaign on outcomes of interest, such as total human capital investment.

A number of other landmark studies of public development programs have utilized the difference-in-difference estimator to rigorously evaluate impacts. The empirical strategy, greatly popularized by Card & Kreuger's (1994) minimum wage study, is an efficient means of estimating treatment effects when randomization or other quasi-experimental methods such as RDD (Regression Discontinuity Design) are not available. Duflo (2001) uses the difference-in-differences estimator to exploit exogenous variation in education access to estimate the impacts of a massive, phased-in, school construction program in Indonesia that jumped from region to region. Habyarimana and Jack (2018) utilize the difference-in-difference estimator to evaluate the impact of a road-safety behavioral intervention in Kenya as a means of controlling for time-invariant pre-intervention differences in group characteristics between treatment and control.

Given the lack of randomization inference regarding India's National Deworming Day, and its status as a large public program, this study will incoporate key insights from relevant deworming, state capacity, development, and public health literature as mentioned above, and take inspiration for its empirical strategy from the quasi-experimental designs that were just reviewed. Additionally, considering the dominance of randomized experiments in the deworming literature, and National Deworming Day's status as one of the world's largest public health campaigns in the world, this paper intends to make a valuable contribution to the literature regarding large scale deworming programs, and the determinants of the efficacy of large governmental programs. Finally, this paper attempts to add more to the evidence base of short-term educational impacts of the deworming treatment, particularly in test scores, in which Miguel & Kremer (2004) failed to find a statistically significant effect.

4 Theoretical Model

This paper will evaluate the impacts of India's National Deworming Day at the national level. Considering public schools as a treatment group and private schools as a control, treatment effects on educational outcomes will be estimated at the national level, in accordance with the size and scale of the program. But while the the study may occur at the national level, microeconomic theory forms the very heart of the theoretical model. Analysis of individual purchasing decisions, productivity, and health and educational outcomes is the key to understanding how the mechanism of the treatment and how one might see impacts at the national level.

This study is highly motivated by empirical results of similar studies in the past, but nonetheless draws inspiration from a number of theoretical models. The most important, and most clearly established model is a simple model of significant positive health externalities associated with treatment. Under any standard economic framework, private consumption of deworming pills is likely to be far below socially optimal levels, and public intervention, including subsidies for the treatment is likely to be welfare-improving across a number of categories, including educational attainment. The merits of this model regarding the deworming treatment are well established in the evidence by Miguel & Kremer (2004, 2019), Ozier (2018) and Bleakley (2007). One should note however, that this literature establishes impacts on educational outcomes by way of gains in equivalent years of schooling. The impacts of the treatment are not analagous to additional education, but rather the recovery or restoration of the impacts of schooling, impacts of schooling that were conceivably lost due to infection. The demonstrated impacts of STH infection could very plausibly lower the return to schooling through their direct impacts on health as well as likelihood to distract and discomfort, sensations that obviously could damage an individual's human capital.

I also consider sub-optimal private consumption of deworming treatments. A number of alternative frameworks are important to consider. First is price elasticity of demand for health products in poor communities explored by Cohen & Dupas's (2010) randomized study of bednet pricing and distribution in Kenya. Cohen & Dupas find that free and subsidized distribution regimes for important, if slightly unpleasant or inconvenient, public health products greatly increase uptake of such products, and by extension, associated health outcomes.

These findings also align well with an S-shaped poverty trap model, in which individuals are implicitly aware of nonlinearities in consumption of such products, and thus are hesitant to spend even small amounts of their limited income on such products. Additionally, an S-shaped poverty trap provides room for imperfect information regarding the role of public health products such as the deworming treatment. If an S-shaped poverty trap does exist, but individuals are not aware of it, they also would likely fail to understand the true benefits of such a treatment. A subsidized distribution regime would be necessary to escape such a trap.

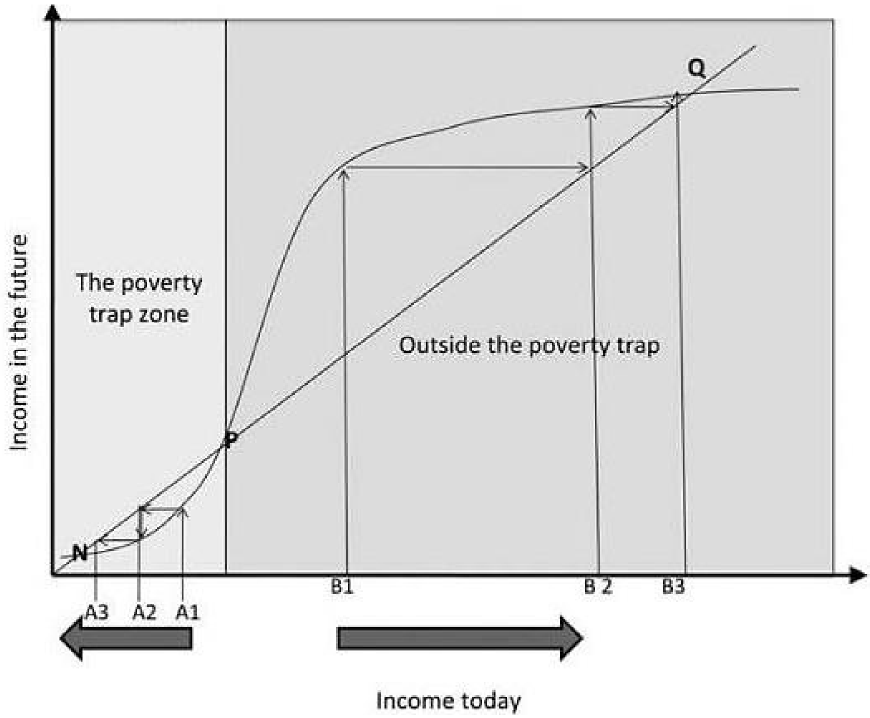


Figure 2: The S-shaped poverty trap (Duflo, Banerjee)

The above model can be thought of as incorporating consumption of nutrition and health goods as investments into productivity, which is transformed into income. Along the same lines, the deworming treatment can be extrapolated as a significant investment in human capital, particularly if it increases school attendance and returns to schooling, as shown by previous studies.

A critical concept to this argument however, is that the upper bound on the treatment effect would be to restore outcomes to the counterfactual scenario, where the children were never infected by Soil-Transmitted Helminths. Then, the treatment effect is a convergence of negatively-impacted students to their better performing peers, rather than a boost in schooling. Following the logic of the Mincer model, where education signals earnings potential, these can be reasonably expected to signal higher future earnings and productivity potential among treated cohorts.

5 Data and Empirical Strategy

5.1 Data

This study uses individual-level survey data of Indian children aged 5-16, and includes measures of cognitive skills, household sanitation levels, and access to healthcare services. Data comes from ASER Centre, an autonomous research and survey organization within the Pratham network in India. ASER Centre conducts annual household and school surveys across India as a part of its Annual Status of Education Report. The ASER survey uses a two-stage sample design to ensure a representative sample of India's rural population. Household level survey data will be used to report cognitive assessment scores and enrollment statistics at the individual (child) level. The cognitive assessments used include reading tests in one of 16 languages (including Hindi, Urdu, English, and Marathi), math tests, and a few English language tests. Each test is graded on a nonlinear basis, using a score of 1 to 5, and measure relatively basic levels of education.

Score	Reading Test	Math Test	English Test
1	Can not read	Can not do arithmetic	Can not read English
2	Can identify letters	Can recognize numbers 1-9	Can identify uppercase English letters
3	Can read words	Can recognize numbers 11-99	Can identify lowercase English letters
4	Can read a level 1 text (pre-primary)	Can do 2-digit subtraction	Can read simple English words
5	Can read a level 2 text (primary)	Can do division (3-by-1 form)	Can read simple English sentences

Table 1: Cognitive Test Scores Dictionary

School enrollment is also recorded by school type, including government schools, private schools, and other miscellaneous schools like Madrasas (Islamic seminary schools). The ASER survey also contains a number of other variables that can be useful in analysis, including household characteristics and access to a number of services within villages. Noting that National Deworming Day 2015 & 2016 targeted government schools nearly exclusively, all students not enrolled in government schools are considered the control group.

The following section provides relevant information on the survey used used for the study. A representative sample was selected from every district in India. Within each district, villages were selected using the probability proportional to size (PPS) sampling method, which properly weights the sample by size and probability of being selected. Thus, the following summary statistics represent a generally accurate description of rural India's population in 2014 and 2016. The data is not a panel data however, as there is limited overlap between the 2014 and 2016 samples. Instead, it is two separate survey samples intended to represent rural India accurately at two distinct time points.

5.2 Descriptive Statistics

First I report descriptive statistics at the national level. Mean levels of educational attainment (following the 1-5 test score method) are clustered around the 3-4 level for the 3 test score variables of interest. Given the nonlinear structure of this data, it is necessary to slice these each of these categories into binary variables in the estimation process, representing attaining a specific test score or higher. Given that most test scores fall around the 3-4 level, I first report descriptive statistics on the perecent of children attaining a score or 4 or higher. The data then shows that attainment of a score of 4 or higher on the cognitive tests remains relatively low, from 25 to 35 percent across all groups.

I also show statistics on household characteristics and village-level amenities. The low prevalence of health clinics lend support to hypotheses that medical treatment is undersupplied, while the lack of household toilet ownership suggests the use of more communal or less hygienic sanitation methods, which implies a high probability of externalities associated with STH transmission. We can also see that general enrollment in government schools, and therefore the treatment group, is around 45 percent of the rural population. The remaining portion of the population is enrolled in private schools, other miscellaneous school types, or not enrolled in school at all.

Table 2: Baseline Characteristics

	Mean	SE	CI
Can at least do two-digit subtraction			
No (n=470,753)	0.741	(0.001)	[0.738,0.743]
Yes (n=172,158)	0.259	(0.001)	[0.257,0.262]
Total (n=642,911)	1.000		
Can at least read a standard level-1 text			
No (n=415,767)	0.649	(0.001)	[0.647,0.652]
Yes (n=227,144)	0.351	(0.001)	[0.348,0.353]
Total (n=642,911)	1.000		
Can at least read simple English words			
No (n=449,368)	0.714	(0.001)	[0.712,0.717]
Yes (n=193,543)	0.286	(0.001)	[0.283,0.288]
Total (n=642,911)	1.000		
Attend Government School			
No (n=347,554)	0.541	(0.002)	[0.539,0.544]
Yes (n=295,357)	0.459	(0.002)	[0.456,0.461]
Total (n=642,911)	1.000		
Attend Private School			
No (n=494,116)	0.777	(0.001)	[0.774,0.780]
Yes (n=148,795)	0.223	(0.001)	[0.220,0.226]
Total (n=642,911)	1.000		
Village has a government clinic			
No (n=364,182)	0.576	(0.005)	[0.567,0.585]
Yes (n=271,490)	0.424	(0.005)	[0.415,0.433]
Total (n=635,672)	1.000		
Village has a private clinic			
No (n=447,824)	0.681	(0.004)	[0.673,0.689]
Yes (n=186,800)	0.319	(0.004)	[0.311,0.327]
Total (n=634,624)	1.000		
Village has an Anganwadi			
No (n=46,939)	0.066	(0.002)	[0.061,0.071]
Yes (n=587,960)	0.934	(0.002)	[0.929,0.939]
Total (n=634,899)	1.000		
Household has a toilet			
No (n=294,310)	0.418	(0.002)	[0.414,0.422]
Yes (n=342,855)	0.582	(0.002)	[0.578,0.586]
Total (n=637,165)	1.000		

Note that sample statistics are calculated using survey-weighted data.

I also provide data on heterogeneity across gender and age groups. A child's age (measured 5-16) is expected to be a significant predictor of their test scores, as older children have access to additional years of education. I show histograms of mean score distributions for each test type by age.

Figure 3: Distributions of Reading Test Scores by Age

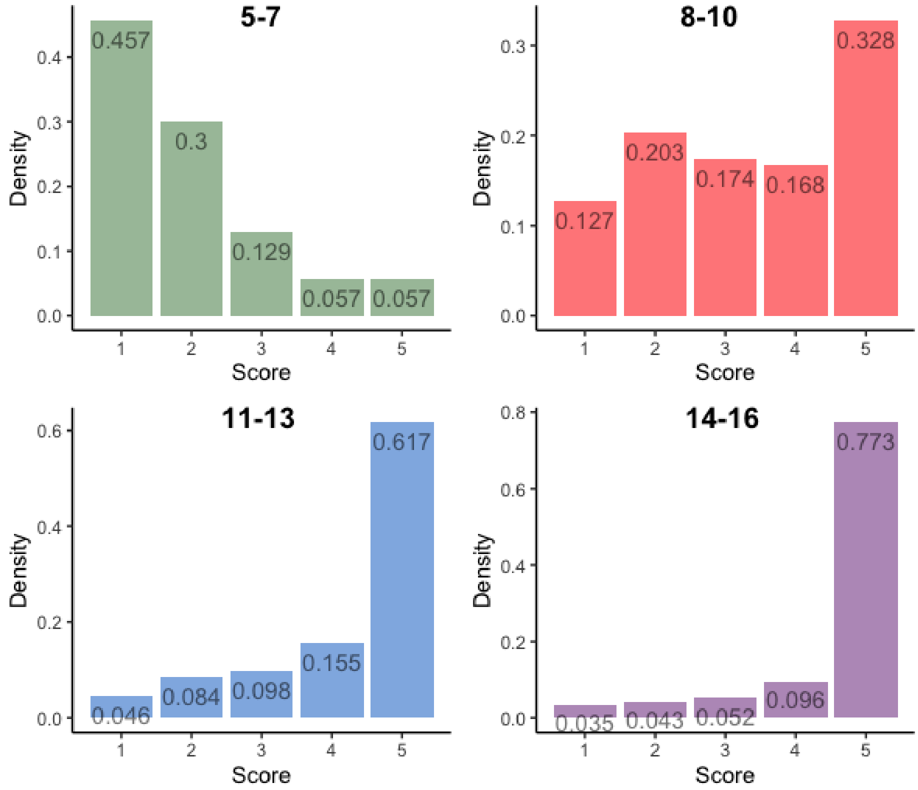


Figure 4: Distributions of Math Test Scores by Age

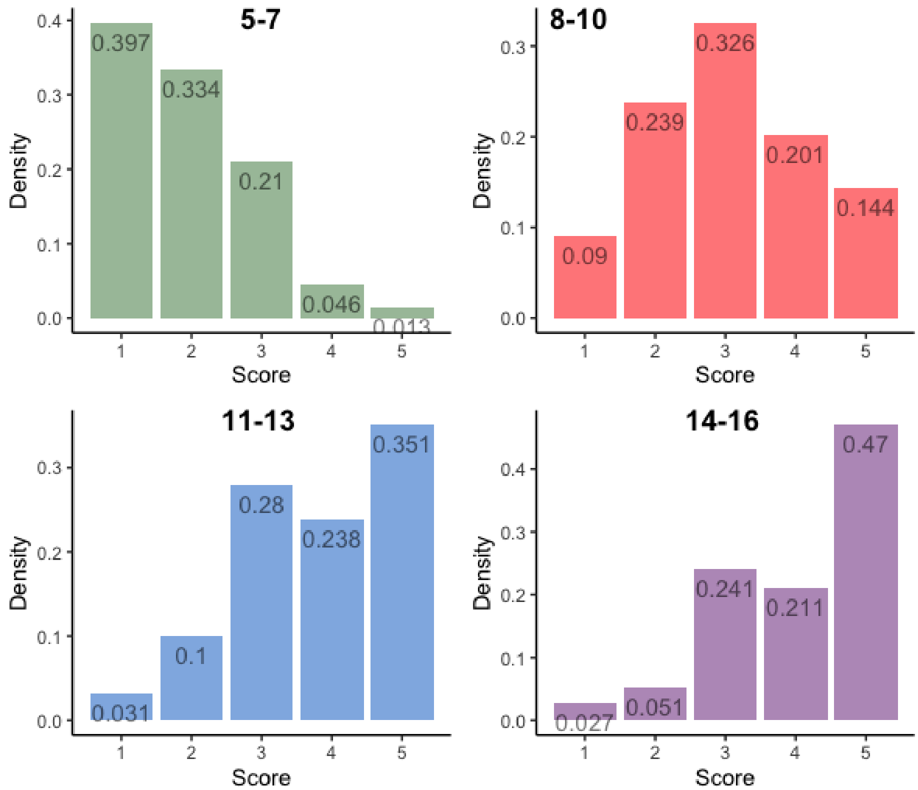
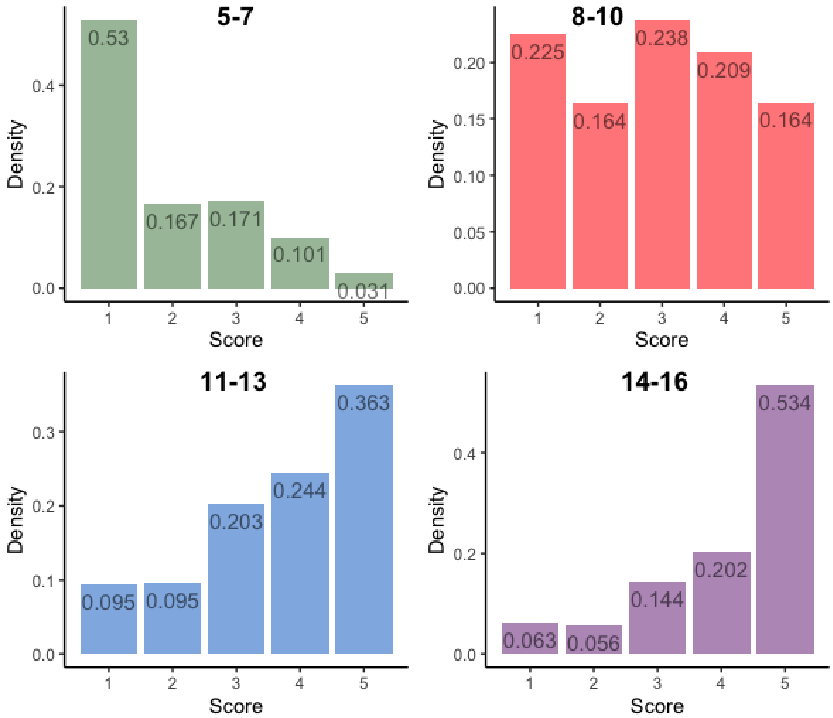


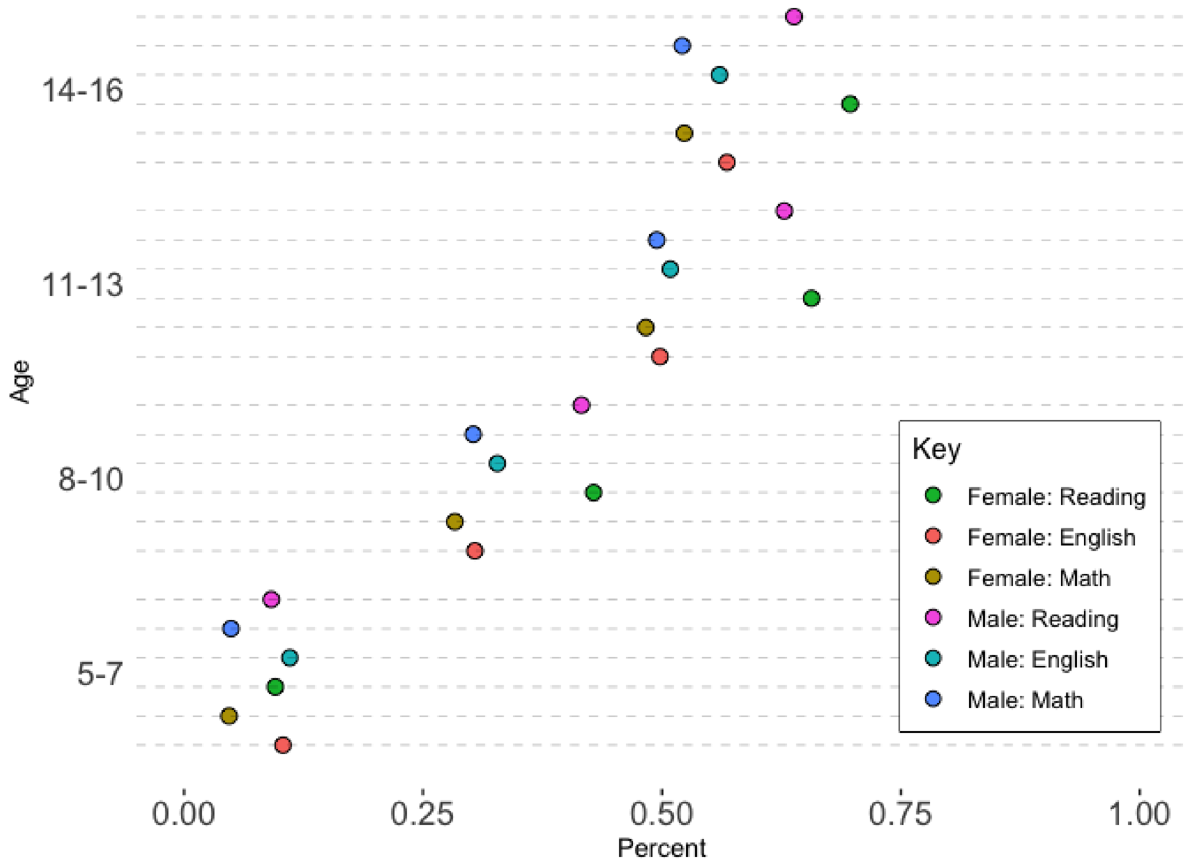
Figure 5: Distributions of English Test Scores by Age



The above distributions show that age is a highly significant predictor of children's test scores, particularly with regards to the reading test. A simple visual analysis of the distributions makes it clear that any expected treatment effect of National Deworming Day will have to be measured differently by age group.

To clarify this finding, the next figure displays attainment of the level-4 benchmark by different age groups. After slicing the test score data into binary categories, I measure the percent of children in age groups that received a score of at least "4" out of 5 on each of the respective cognitive tests. The math benchmark measures whether a child can at least do two-digit subtraction, the reading benchmark measures whether a child can read a standard pre-primary school level text, and the English benchmark measures whether a child can read simple English words. One can see that test scores, as expected, clearly increase with age. Furthermore, one can see that approximately 50 percent of children attain the level-4 benchmark around the 11-13 age group, across test type and gender. For younger children however, attainment is incredibly low — below 15%. Accordingly, I define age-wise attainment benchmarks for each age group. I estimate treatment effects for the probability of obtaining a score of at least "3" for children aged 5-10, and the probability of obtaining a score at least "4" for children aged 11-16.

Figure 6: Percent of children reaching at least level-4 benchmark learning goals



5.3 Balance Test

I also conduct a balance test between publicly and privately schooled children at baseline (2014) and find that highly statistically significant differences exist between the two groups across a wide range of variables. Significant baseline differences were expected however, which is why the study employs a difference-in-differences (DiD) design. The specifications and qualifying assumptions for the DiD design are discussed below.

Table 3: Balance Test

	(1)		(2)		T-test
	0		1		Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Can at least do two-digit subtraction	303157	0.437	588348	0.345	0.092***
		(0.001)		(0.001)	
Can at least read a standard level-1 text	303157	0.541	588348	0.467	0.074***
		(0.001)		(0.001)	
Can at least read simple English words	303157	0.534	588348	0.364	0.169***
		(0.001)		(0.001)	
Village has a government clinic	299959	0.472	582535	0.397	0.075***
		(0.001)		(0.001)	
Village has a private clinic	299739	0.349	581963	0.244	0.105***
		(0.001)		(0.001)	
Village has an internet cafe	299821	0.241	582117	0.151	0.090***
		(0.001)		(0.000)	
Village has an Anganwadi	299434	0.917	581803	0.926	-0.009***
		(0.001)		(0.000)	
Household has a toilet	301141	0.350	583510	0.576	-0.226***
		(0.001)		(0.001)	

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

6 Empirical Strategy

6.1 Econometric Model

The study uses a difference-in-differences framework, where students attending government schools are considered the quasi-experimental treatment group and post is a dummy variable, 1 for data collected in 2016, 0 for data collected in 2014. Given the nonlinear, yet ordinal nature of the outcome variables, I consider a number of alternative regression specifications. Relying on past literature and ease of interpretation, I first use a binary response variable[2] with a double-difference framework within a linear probability model estimated by OLS (Ordinary Least Squares). I then compare results and perform a robustness check using a logit model. I then explore heterogeneous effects and the impact of state capacity using a triple-difference framework, again using a linear probability model estimated by OLS, and performing robustness checks using a logit model. I then explore modelling the ordinary response variable using an Generalized Ordered Logistic Model, where I

[2]Whether the child received a score of at least 3 for children aged 5-10, at least 4 for children 11-16.

measure the odds-ratios of the treatment group falling into each response category. Because the application of double and triple-difference specifications in ordered logit models is under-explored in econometrics literature, I emphasize the results of the linear probability model, which has been repeatedly shown to reliably estimate treatment effects under generalized DiD specifications.

To present intial results, I use a linear probability model as well as a logit model to estimate the treatment's effect on the probability of students obtaining the selected testing benchmark for their age group, or higher, out of 5. The linear probability model, estimated by OLS, uses the following form.

$$\hat{Y} = \hat{\alpha} + \hat{\beta}(Treatment) + \hat{\omega}(Post) + \hat{\delta}(Treat * Post) + \Sigma_i^N \gamma_i(X_i) \tag{1}$$

Where Y refers to the overall mean of the group by treatment status (publicly or privately schooled Indian children) on a given variable of interest. α is the constant term, β captures pre-existing differences between the two groups, ω is the effect of the time trend from 2014 to 2016 as national data collection was paused in 2015, and γ_i is the set of coefficients for a vector of control variables, X_i. The final, interaction term, δ captures the difference-in-differences.

I include a number of control variables to increase the precision of the estimates, including state-level effects and dummies for gender, household ownership of a toilet, and the local presence of government and private health clinics. Family income is not available in the survey data.

I next consider a standard logit model, where the outcome variable is the log odds ratio of at least obtaining the selected benchmark score, and the independent variables follow from the Linear Probability Model.

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \hat{\alpha} + \hat{\beta}(Treatment) + \hat{\omega}(Post) + \hat{\delta}(Treat * Post) + \Sigma_i^N \gamma_i(X_i) \tag{2}$$

Where the left side represents the log odds of the dichotomous outcome, (receiving a score equal to or greater than the selected benchmark), and the right side follows the same as the above model. The difference is the coefficients in this form represent changes in log-odds ratios of obtaining the outcome. In order to report changes in probability, one can take the coefficients as follows.

Then $\hat{\pi}$ is the predicted probability of obtaining a test score of at the benchmark or higher,

$$\hat{\pi} = \frac{e^{\alpha+\beta(Treat)+\omega(Post)+\delta(Treat*Post+\Sigma_i^N \gamma_i(X_i)}}{1+e^{\alpha+\beta(Treat)+\omega(Post)+\delta(Treat*Post+\Sigma_i^N \gamma_i(X_i)}} \tag{3}$$

To obtain more precise results and estimate differential impacts by varying levels of State capacity, I also employ a triple-difference estimator. In the triple difference framework, I define non-publicly schooled students as a standard control group, or the first difference, and states that were only treated in 2016 as the second difference.The third difference is then private school children who were in high-capacity states. Further explanation of the motivation and inference follows in the next section. The triple difference estimator takes the form as follows

$$\hat{Y} = \hat{\beta_0}+\hat{\beta_1}(T)+\hat{\beta_2}(P)+\hat{\beta_3}(N)+\hat{\delta_0}(T*P)+\hat{\delta_1}(T*H)+\hat{\delta_2}(P*H)+\hat{\delta_3}(T*P*H)+\Sigma_i^N\gamma_i(X_i) \quad (4)$$

Where T refers to children enrolled in public-schools, P refers to data collected in 2016, and H refers to high-capacity states that were treated in 2015 in addition to 2016. The primary coefficient of interest in this regression is δ_3, on the triple interaction term, $T * P * H$. The δ_3 coefficient can similarly be expressed in a logistic regression model as follows:

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \hat{\beta_0}+\hat{\beta_1}(T)+\hat{\beta_2}(P)+\hat{\beta_3}(N)+\hat{\delta_0}(T*P)+\hat{\delta_1}(T*H)+\hat{\delta_2}(P*H)+\hat{\delta_3}(T*P*H)+\Sigma_i^N\gamma_i(X_i) \quad (5)$$

Next, I attempt to model for the ordinal structure of the test score data (scored 1-5). To do so, I present equations for the Generalized Ordered Logistic Regression model, which estimates the log odds ratios (or probabilities) of falling into each response category, under both the double and triple difference framework. I use the Generalized Ordered Logit Model (GOLOGIT) opposed to the Proportional Odds Model to account for violations of the proportional odds assumption.[3] To account for this violation, I estimate 5 equations, one for each response category. I present the equations and results from the equations in terms of probability, following the method used in equation (5).

$$\hat{P}(Y > j) = \frac{e^{\hat{\alpha_j}+\hat{\beta_j}(T)+\hat{\omega_j}(P)+\hat{\delta_{ij}}(T*P)+\Sigma_i^N\hat{\gamma_j}(X_i)}}{1+e^{\hat{\alpha_j}+\hat{\beta_j}(T)+\hat{\omega_j}(P)+\hat{\delta_j}(T*P)+\Sigma_i^N\hat{\gamma_{ij}}(X_i)}}, \quad j = 1, 2, ..., M-1 \quad (6)$$

Where $\hat{P}(Y > j)$ refers to the estimated probability of a child's test score falling into a category j, when Y has j+1 categories. α_j, β_j, ω_j & δ_{ij} refer to the coefficients for response category j or lower relative to all categories greater than j. Under the Proportional Odds Model, one would assume that only the constant term, α, would be indexed by j.

The reader should note that estimation of difference-in-difference coefficients within Ordered Logit Models is a relatively untested empirical strategy, and the coefficients are difficult to interpret. The results of the Ordered Logit Models should be interpreted with restraint, although they provide a useful means of modelling the ordinal response and act as an additional robustness check on preceding regression specifications.

Due to the differing levels of attainment among age groups, and to explore differences by gender, I stratify the sample by gender and age of the children. Stratifying on these lines is important as potentially significant heterogeneous effects on both gender and

[3]That the coefficients for each independent variable are equal across response categories.

age lines have been documented. Miguel & Kremer (2004) note that many practitioners advise not providing deworming treatment to girls of reproductive age, due to concern about the possibility of birth defects. Although no evidence of this practice was found in India's National Deworming Day, I consider it nonetheless. Then, proper stratification will allow a comparison of girls of reproductive age both with boys or girls of younger ages. Additionally, given the differences between age groups, stratification is likely to increase the statistical power of the test.

Standard errors in the regression are also clustered at the village level. Following Miguel & Kremer (2004, 2019) and Bleakley (2007), clustering of errors is an essential step when estimating the impact of a deworming treatment. While Miguel & Kremer cluster at the school level in their studies, Bleakley clusters on state economic areas due to the nature of data availability. Following Miguel & Kremer (2004)'s discovery of treatment externalities beyond the directly treated pupils in attendance, as well as data limitations, I cluster standard errors at the village level times school type, that is one cluster for public schools and one for private schools, each at the village level.

6.2 Difference-in-Differences Design

The difference-in-differences design rests primarily on the fairly strong assumption of parallel trends between treatment and control groups in the time in which baseline statistics are calculated and impact evaluation estimates are made. In the context of this study, parallel trends are threatened by a neoclassical convergence effect, in which lower-scoring children in government schools converge to the level of private school children more quickly. A test of preexisting trends prior to the treatment will explore this possibility. Perhaps more plausibly, parallel trends are likely to be threatened by the effects of alternative programs during the study period. In simple words, although we may detect a significant difference-in-differences between the two groups under this design, it may be difficult to verify that the difference can be attributed to National Deworming Day, rather than alternative programs nationally that are correlated with treatment status. An analysis of pre-intervention trends follows in the results section.

Another meaningful challenge to the validity of the study is the Stable Unit Treatment Value Assumption (SUTVA). The first component of the SUTVA assumption requires that potential outcomes in relation to treatment are well defined, that is, multiple levels of treatment cannot exist. While this may be difficult to satisfy at the individual level, where recieving multiple deworming pills regularly is certainly a "better treatment" than just receiving one, this assumption is likely to hold under the study design for two reasons. Firstly, National Deworming Day is defined by its status as a single day program, therefore there is likely to be no differences in the level of treatment; any differences in the total amount of deworming treatments taken is not likely to be endogenous within the context of the intervention. However this cannot be verified — due to a lack of data availability, I am not able to reject the possibility that the treatment and control groups have differential access to treatment outside the scope of national deworming day.

Secondly, the SUTVA assumption requires a sufficient handling of spillover effects within treatment. A naive analysis that fails to account for spillover effects is certain to produce a biased treatment effect estimate. As mentioned earlier, I adjust for the certainty of spillover effects by clustering standard errors at the village level, consistent with the literature's demonstration of externalities associated with the treatment, and the possibility of spillovers to control groups within villages. I am unable, however, to fully account for possibility of treatment spillovers to the control group, which is likely to bias the treatment effect estimates towards zero.

As mentioned above, I also generalize the difference-in-difference estimator to a triple-difference. Noting that 12 of 36 states and union territories were treated in 2015 in addition to 2016, and that these states were selected by the Ministry of Health and Family Welfare by virtue of "the States preparedness for effectively conducting the deworming round," it becomes important to consider these 12 states as significantly differing from the other 24 for a number of reasons. Firstly, given the spread of data collection by the ASER survey in 2016, the impacts of the 2016 treatment may not be properly captured. More significantly, the selection of the 12 states for an initial treatment in 2015 creates an opening for a method of measuring differences in outcomes between high capacity and low capacity states. The triple-differences coefficient, then, represents the differences in the impact of the treatment between high-capacity and low-capacity states, and allows me to quantify impacts of high levels of state capacity on administering government public health programs effectively.

Mechanically, the triple-difference estimate rests on a less strict assumption, the Uniform Convergence Assumption. Rather than requiring parallel trends between publicly and privately schooled students, the triple-difference requires that the rate of change, or second order difference between private and publicly-schooled children, to be uniform over time. This assumption is relaxed in comparison to the simple difference-in-differences design, which requires equivalent first-order differences between treatment and control groups. In this way, the key assumption of the triple difference model is the lack of additional shocks that are endogenous with treatment status.

6.3 Compliance, Sampling and Power

Following this approach, I employ a standard Intention-to-Treat (ITT) estimation strategy, in which all public school children are assigned treatment status, which has a number of advantages. First of which is lack of reliable data on which schools were treated, and to that extent, making a Treatment-on-the-Treated (ToT) strategy nearly impossible. Secondly, per Miguel & Kremer (2004), an ITT strategy is an effective means of addressing potential transfer bias, such as students switching school enrollment. Primarily however, an ITT strategy also best fits the objective of the study — to evaluate the impact of a national public program on national outcomes; to determine the impact of National Deworming Day 2015 and 2016 on a representative national sample. According to Evidence Action estimates, 89 million children were treated in 2015, followed by 179 million children in 2016, totalling 268 million treated children. Conservative estimates place the number of children in India from 2015-2016 at around 470 million, implying that an ITT framework is the most appropriate means of evaluating the program's impacts on the country as a whole.

The sheer scale of the program also reduces fears of an under-powered study following an ITT strategy. A naive analysis would conclude that as many as 60% of the country's children were treated by the program, and although the true number is nearly certain to be lower, on account of a compliance issue and the possibility of individual subjects being "double-counted" in treatment, it is clear that the outreach of the program was extremely high. As mentioned above, one disadvantage of the deworming treatment is that it needs to be taken regularly, meaning the number of unique children receiving the treatment could be significantly lower than mentioned above. Alternatively, given the nature of the infection and need for regular treatment, it may be useful to view successive applications to the same individual as two distinct treatments. For these complications, a simple Intention-to-Treat strategy provides the most straightforward estimation method.

The study, like nearly all public program evaluations, also suffers from a minor compliance issue. While National Deworming Day 2015 targeted exclusively government schools, NDD 2016 included a small number of private schools by design. Of the 1.79 million treated schools in 2016, 170,000, or just under 10%, were private schools. Data limitations do not allow an estimation of the compliance rate at the individual level, so the true compliance rate is unknown, but it is all but certain that an ITT strategy will somewhat underestimate the impact of treatment, as a number of control observations did receive it.

Given the study's limitations in estimation, a. careful consideration of statistical power is needed. National level studies are always prone to issues of insufficient power. India's National Deworming Day however, is unique in the scope of its outreach. 89 million, then followed by 179 million children in India were treated in 2015 and 2016, according to Evidence Action. Noting that the total population of the country is estimated at 1.36 billion in 2019, I hypothesize that the outreach of the program is sufficiently high to expect the possibility of a detectable effect.

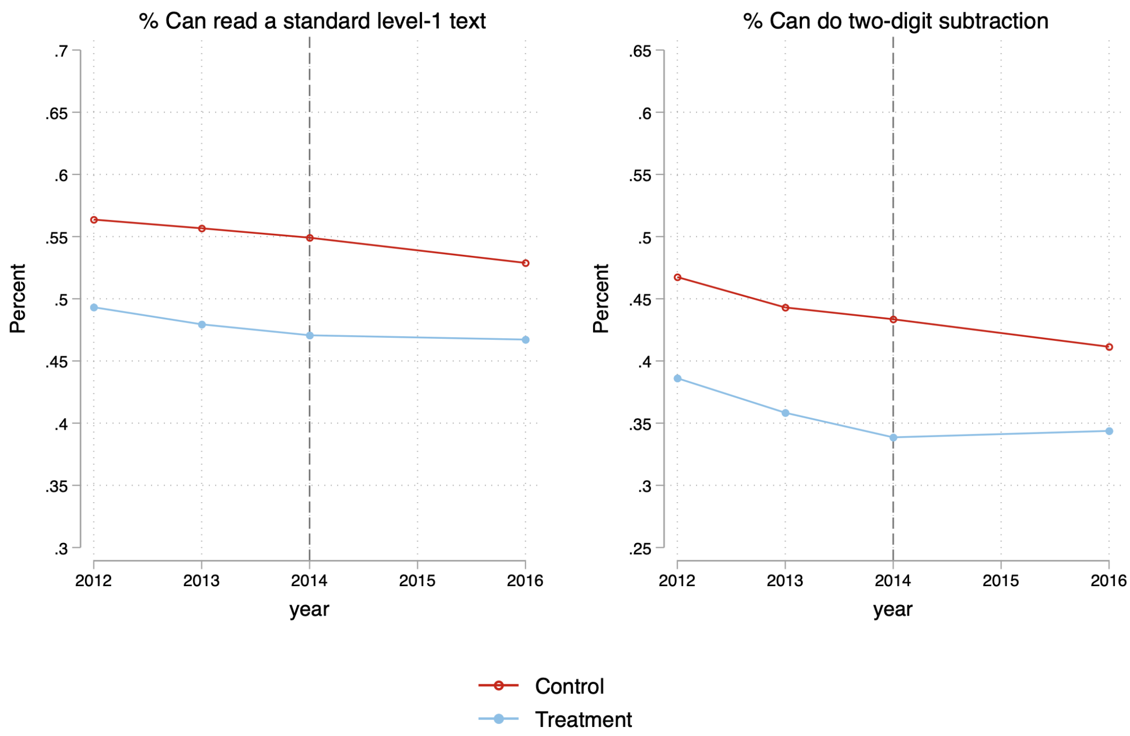
7 Results

Here I present estimated treatment effects on Math, Reading, and English test scores on rural Indian children aged 5-16 attending government schools. I first present results in terms of the probability each age group attaining at least the benchmark score selected above for both the linear probability and logistic model. Following discussion of primary results, I also further explore heterogeneous effects by age, gender, and high-capacity states. Then, I present results using an ordered logit model, for both double and triple difference specifications. All regression estimates are weighted according to the specifications of the ASER survey, and standard errors are clustered at the village level. Before I present results however, I provide an analysis of the parallel trends assumption.

7.1 Parallel Trends

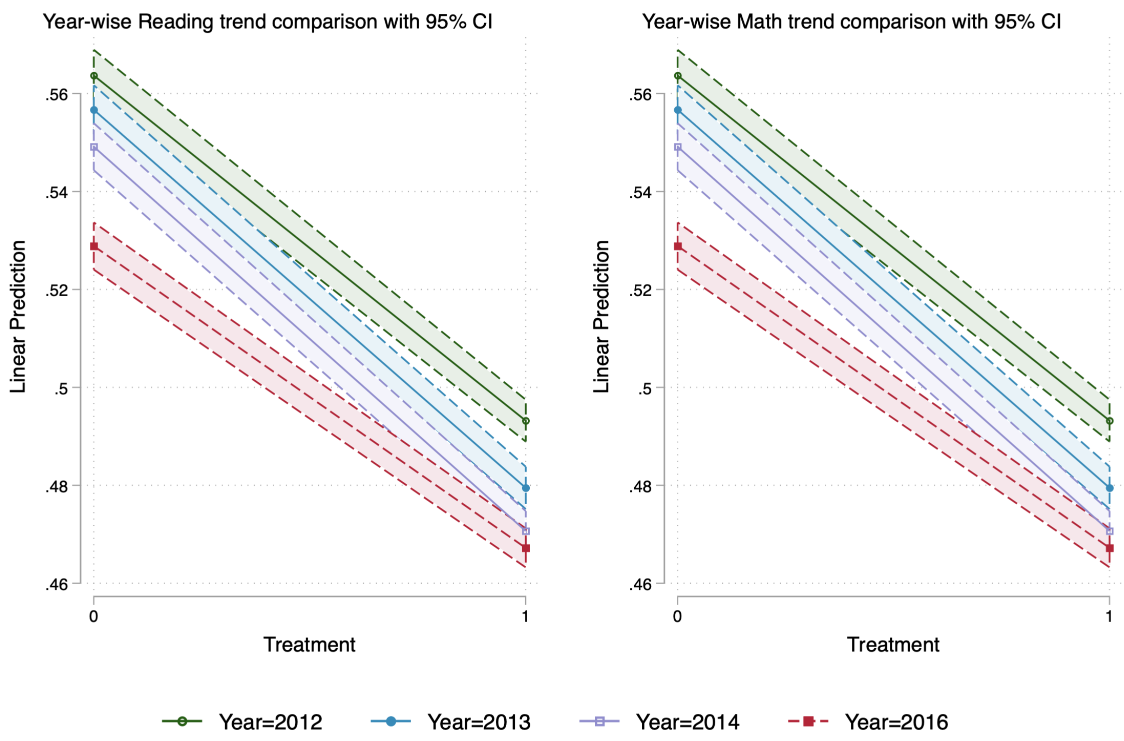
Figure 8 shows that the preexisting trends appear to be parallel from 2012-2014, the 3 years prior to the treatment. I compare trends in math and reading scores, as estimates of English test scores prior to the treatment period are not as reliable. The figure visually compares trends between the treatment and control groups, and shows a parallel structure. One can notice minor discrepancies at the 2013 point estimate, but a general parallel trend is observed between the two groups in both math and reading scores. One challenge to further establishing the validity of the assumption is the lack of data in 2015, but regardless one can clearly see a divergence between the two groups from 2014 to 2016

Figure 7: Pre-treatment Trends



For further comparison, I plot year-over-year differences between treatment and control, in the years of interest in Figure 9. Each line corresponds to the difference between groups in a given year, with a set of parallel lines representing parallel trends between two time periods. The green, blue and purple lines represent pre-treatment years, and the red line represents 2016 (post-treatment) for comparison. In both cases, one can see a nearly parallel structure in the pre-treatment years, particularly in the the 2 years immediately preceding the treatment.

Figure 8: Trend Comparisons with confidence intervals



7.2 Primary Results

I find that India's National Deworming Day had significant effects on Reading, Math, and English Test Scores for children aged 11 or older. I estimate a difference-in-differences coefficient using both a Linear Probability and Logit Model, controlling for a number of variables recorded at the household level. I include dummy variables for each state, as well controls for whether or not an individual has a toilet in their household, has a private or government health clinic in their village, or an Anganwadi (a government-run childcare center) in their village.

Using a Linear Probability Model, I find that, across India, National Deworming Day increased the probability of rural Indian children aged 11-16 attending government school of successfully reading a standard level-1 text by just over 1 percentage point. I find even larger effects for Math and English scores, I estimate that the program increased the probability of the sample children being able to successfully perform two-digit subtraction and successfully read simple English words by approximately 2.4 percentage points. Change-in-probability estimates from the logistic model are very similar, and can be seen below in Table 4. I report coefficients on treatment, referring to government schools, and post, referring to data from the 2016 National ASER survey in addition to the estimate treatment effect from the Linear Probability Model. I also report the estimated treatment effect for the logit model.[4]

[4]STATA code and inference were provided by Dr. Jeffrey Woolridge via #EconTwitter: https://twitter.com/jmwooldridge/status/1366033635191701508?s=20

Table 4: Primary Results

	Linear Model			Logit Model		
	Reading	Math	English	Reading	Math	English
5-10: Treat	-0.0968***	-0.1364***	-0.2126***			
	(0.0036)	(0.0036)	(0.0037)			
5-10: Post	-0.0104**	-0.0101**	-0.0064			
	(0.0044)	(0.0044)	(0.0045)			
5-10: Treat*Post	0.0045	0.0070	0.0158***	0.0043	0.0070	0.0166***
	(0.0050)	(0.0049)	(0.0050)	(0.0049)	(0.0048)	(0.0048)
5-10: β_0	0.6412***	0.7707***	0.7906***			
	(0.0097)	(0.0094)	(0.0096)			
11-16: Treat	-0.0461***	-0.0931***	-0.1213***			
	(0.0035)	(0.0039)	(0.0038)			
11-16: Post	-0.0326***	-0.0435***	-0.0409***			
	(0.0042)	(0.0048)	(0.0046)			
11-16: Treat*Post	0.0105**	0.0236***	0.0236***	0.0140***	0.0245***	0.0264***
	(0.0049)	(0.0054)	(0.0052)	(0.0053)	(0.0055)	(0.0055)
11-16: β_0	0.8243***	0.7901***	0.8498***			
	(0.0095)	(0.0105)	(0.0101)			

Note: Controls include state, presence of clinics and Anganwadis, and household toilet access.
Sample sizes are 439,602 and 425,745 for each age group, respectively.

In Figures 10 and 11 I also provide visual displays of treatment effects for the two models, with confidence intervals displayed at the 95% (vertical bars) and 99% (tails of the horizontal line) levels around the mean estimates. One can see that estimates are significant at the 95% level for all outcomes for children aged 11-16, and significant at the 99% level for Math and English scores. Although neither Reading or Math Scores are significant at the 95% level for children-aged 5-10, English scores are significant at the 99% level. One can also see that the magnitude of most of the coefficients increases slightly under the logistic regression model.

The heterogenous effects by age and test type are of particular importance. For example, the findings suggest that the first two years of India's National Deworming Day increased the amount of rural Indian children aged 11-16 who can perform two-digit subtraction by as much as 2.5 percentage points nationwide, while it only increased the same group's ability to successfully read a standard level-1 text by a little over 1 percentage point. For children aged 5-10, National Deworming Day was not found to have any significant effects on their ability to successfully read fully-formed words in their native language or recognize numbers 11-99. It was however, found to have a significant impact on their ability to identify lowercase English letters, improving the number of children who could by over 1.5 percentage points.

Figure 9: Treatment Effects: Linear Probability Model

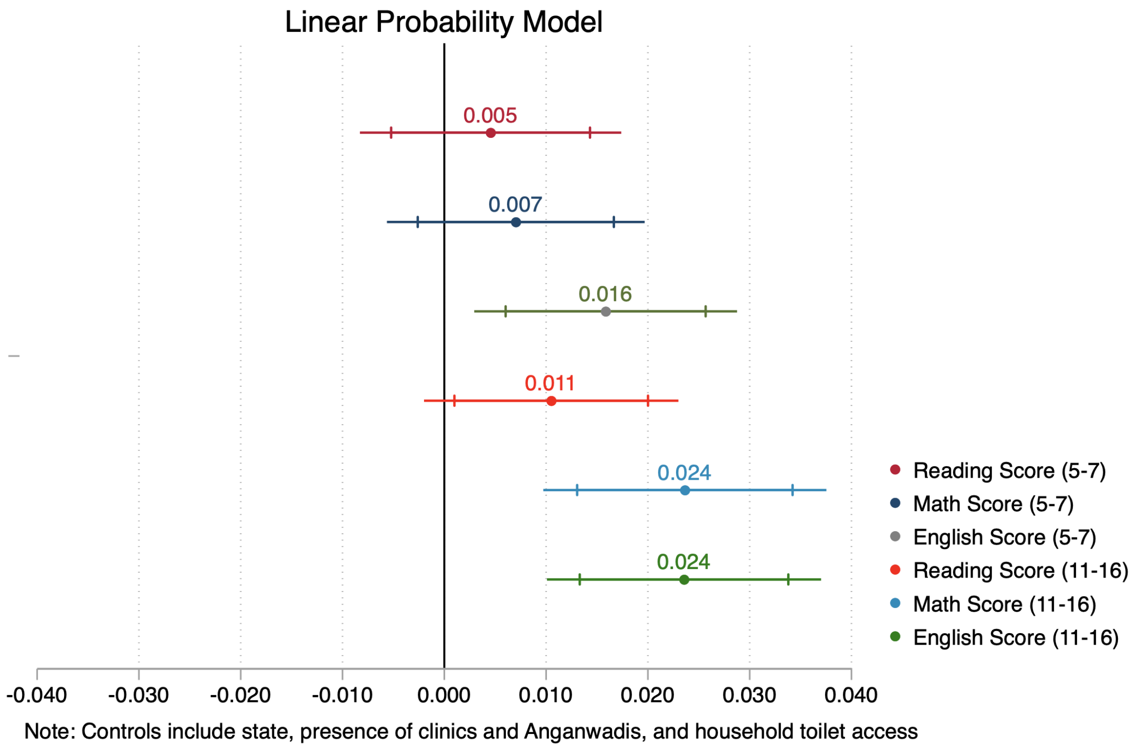
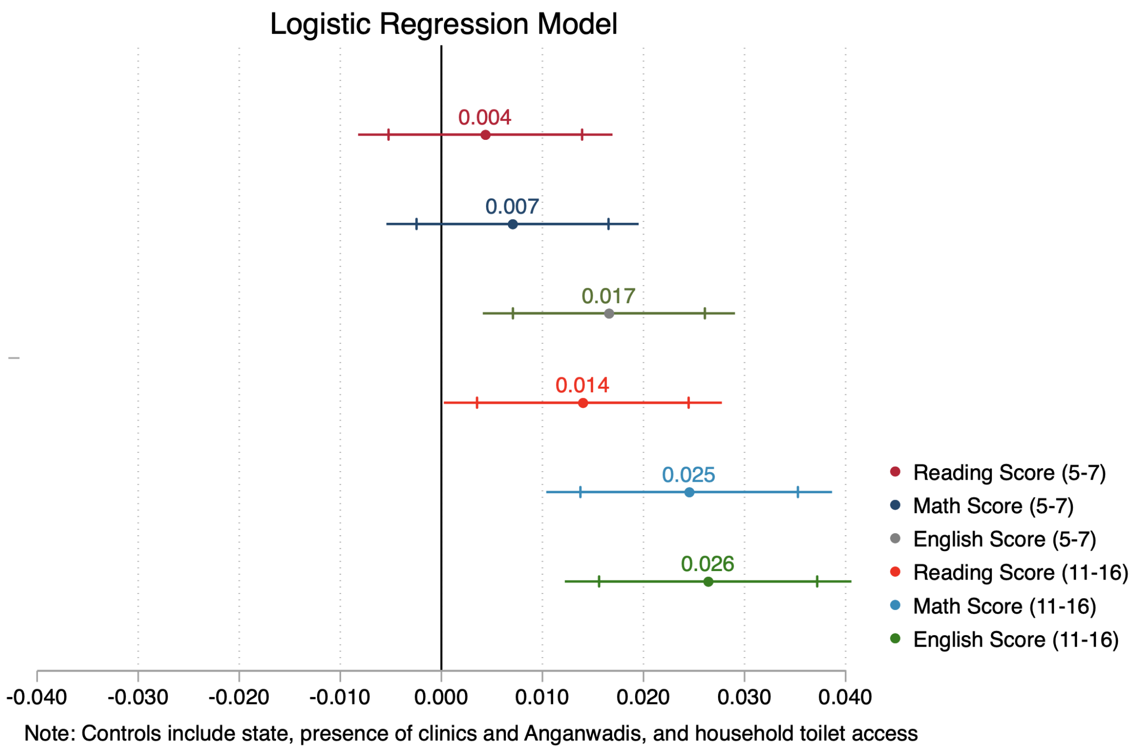


Figure 10: Treatment Effects: Logistic Regression Model

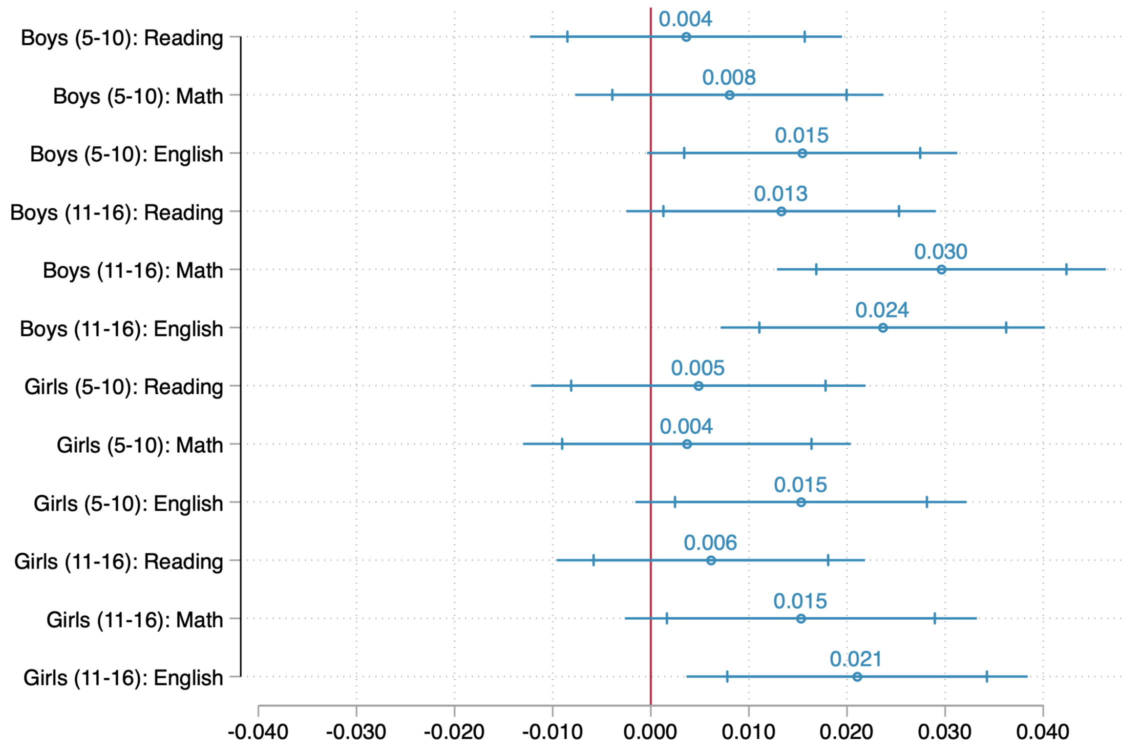


7.3 Heterogeneous Treatment Effects

I continue to explore heterogeneity in the results. Firstly, I further explore heterogeneity by age, including results for alternative benchmarks for younger age groups. Next I analyze results by gender and state, noting baseline differences in gender as well as differences in treatment status by state.[5]

First, I plot regression coefficients along gender and age lines for visual comparison by the reader.

Figure 11: Heterogeneous Treatment Effects: Gender



From visual inspection, one can see estimates for girls are mostly insignificant and of smaller magnitude, implying that the treatment was less impactful for girls. One dimension of interest briefly discussed earlier in the paper was the possibility of girls of child-bearing age not being given the treatment, as is mentioned by Miguel & Kremer et al. (2004). As previously noted, I found no evidence of India's National Deworming Day making this treatment exclusion, and the estimated treatment effects for girls of this age are robust. In fact, the estimated treatment effects for girls of child-bearing age are of a much greater magnitude than the effects of children of lower ages, a result that would be inconsistent with girls of child-bearing age not receiving the treatment.

[5]I herein refer to the different groups in the survey sample as "boys" and "girls" referring to the recording of their gender by the ASER survey. The survey uses a simple binary code to indicate if the child is female or male. I do not intend to make any assumptions about the childrens' preferred gender identity.

The treatment effects for girls aged 11-16 is less statistically significant and of approximately half magnitude for reading and math scores. While boys are estimated to have gained a 3 percentage point increase in their probability of successfully performing two digit subtraction and a 1.3 percentage point increase in their probability of successfully reading a standard level-1 text, the corresponding effects for girls are just 1.5 and .6 percentage points, respectively. The estimated effect for reading scores for girls is also not statistically different from zero. In English scores however, the estimated treatment effects for boys and girls are 2.4 and 2.1 percentage points respectively, and both are significant at the 99 percent level. Despite the differences in the magnitude of treatment effects for boys and girls however, I am not able to reject the null hypothesis that girls were not impacted differently by the treatment. Results are shown in Table 5.

Table 5: Heterogeneous Treatment Effects: Gender

	5-7			11-16		
	Reading	Math	English	Reading	Math	English
Treat	-0.0949***	-0.1331***	-0.2091***	-0.0457***	-0.0912***	-0.1160***
	(0.0044)	(0.0044)	(0.0045)	(0.0043)	(0.0047)	(0.0046)
Post	-0.0128**	-0.0142***	-0.0072	-0.0358***	-0.0485***	-0.0458***
	(0.0052)	(0.0051)	(0.0052)	(0.0051)	(0.0056)	(0.0054)
Female	0.0166***	-0.0137***	-0.0001	0.0575***	0.0072	0.0174***
	(0.0046)	(0.0045)	(0.0045)	(0.0041)	(0.0046)	(0.0044)
Treat*Post	0.0029	0.0073	0.0145**	0.0130**	0.0282***	0.0231***
	(0.0062)	(0.0061)	(0.0061)	(0.0061)	(0.0065)	(0.0064)
Female*Post	0.0057	0.0108*	0.0031	0.0081	0.0117*	0.0118*
	(0.0065)	(0.0063)	(0.0062)	(0.0059)	(0.0065)	(0.0064)
Female*Treat	-0.0070	-0.0040	-0.0069	-0.0097*	-0.0047	-0.0132**
	(0.0056)	(0.0055)	(0.0055)	(0.0052)	(0.0056)	(0.0056)
Treat*Post*Female	0.0026	-0.0030	0.0018	-0.0063	-0.0110	-0.0011
	(0.0079)	(0.0078)	(0.0077)	(0.0074)	(0.0080)	(0.0080)
β_0	0.6345***	0.7770***	0.7905***	0.7991***	0.7873***	0.8426***
	(0.0099)	(0.0096)	(0.0098)	(0.0097)	(0.0108)	(0.0103)
Number of Observations	436259	436259	436259	422915	422915	422915

Note: Controls include state, presence of clinics and Anganwadis, and household toilet access

Another key dimension of heterogeneity comes from States. I estimate differential treatment effects for states that were designated by the Ministry of Health and Family Welfare as having high levels of capacity to administer the treatment widely. I present results from the linear probability model's triple-difference estimation, using publicly-schooled children as a treatment, and states that were not designated as high capacity as a synthetic control for the triple difference. Results are shown in Table 6. Coefficient labels follow as before, but high capacity states are represented by "NDD2015".

Table 6: Results for High-Capacity States (Triple-Difference)

	5-7			11-16		
	Reading	Math	English	Reading	Math	English
Treat	-0.1092***	-0.1395***	-0.2110***	-0.0607***	-0.0988***	-0.1159***
	(0.0050)	(0.0049)	(0.0051)	(0.0051)	(0.0056)	(0.0055)
Post	-0.0144**	-0.0199***	-0.0160***	-0.0391***	-0.0636***	-0.0539***
	(0.0058)	(0.0058)	(0.0058)	(0.0058)	(0.0066)	(0.0063)
NDD2015	0.0011	-0.0836***	0.0406**	-0.0533***	-0.1218***	-0.0443**
	(0.0206)	(0.0207)	(0.0200)	(0.0183)	(0.0222)	(0.0204)
Treat*Post	-0.0148**	-0.0091	0.0076	-0.0059	-0.0024	0.0019
	(0.0071)	(0.0070)	(0.0073)	(0.0072)	(0.0079)	(0.0077)
NDD2015*Post	0.0134	0.0265***	0.0250***	0.0170**	0.0482***	0.0305***
	(0.0087)	(0.0087)	(0.0090)	(0.0082)	(0.0096)	(0.0091)
NDD2015*Treat	0.0348***	0.0162**	0.0035	0.0356***	0.0238***	-0.0027
	(0.0072)	(0.0071)	(0.0074)	(0.0069)	(0.0078)	(0.0075)
Treat*Post*NDD2015	0.0247**	0.0163	0.0047	0.0202**	0.0279**	0.0269**
	(0.0100)	(0.0099)	(0.0102)	(0.0097)	(0.0109)	(0.0105)
β_0	0.6537***	0.7794***	0.7948***	0.8404***	0.8077***	0.8562***
	(0.0098)	(0.0096)	(0.0097)	(0.0098)	(0.0108)	(0.0104)
Number of Observations	439602	439602	439602	425745	425745	425745

Note: Controls include state, presence of clinics and Anganwadis, and household toilet access.

The triple difference estimator enjoys increased precision and relaxed assumptions relative to the double-difference estimator. From the results, one can see clearly that the high capacity states [6] were the epicenter of National Deworming Day's impact in 2015 and 2016. While the estimated treatment effects for the high capacity states are of marginally higher magnitude relative to other states, the estimated treatment effects for states not designated as high capacity are also not statistically different from zero. It should be noted that estimates for Math and English scores for children aged 5-10 are not significant, but that treatment effects for Reading, and Math and English Scores for children aged 11-16 are statistically significant at the 95% level, and results for Math scores 5-10 are marginally below the 10% level of significance. More precisely, the first two years of National Deworming Day increased the probability of rural children 11-16 successfully performing two digit subtraction by nearly 2.8 percentage points, a significant finding given the scope of the program and study. Other estimated treatment effects within high capacity states are 2.47 percentage points in reading fully-formed words for children aged 5-10, 2 percentage points in reading standard

[6] Assam, Bihar, Chattisgarh, Dadra and Nagar Haveli, Haryana, Karnataka, Maharashtra, Madhya Pradesh, Rajasthan, Tamil Nadu, and Tripura.

level-1 texts for children aged 11-16, and 2.69 percentage points in reading simple English sentences for children aged 11-16.

These findings present two different interpretations. First, they demonstrate that the states designated as prepared to carry out the deworming process effectively by the Ministry of Health and Family Welfare performed significantly better in administering the deworming treatment widely to their rural population. In fact, these findings suggest that the capacity and infrastructure of these states are necessary prerequisites to make such a public health intervention successful at all. They also imply that local governments have the ability to self-assess their capacity to carry out large scale programs.

However, it is likely that the results contain minor structural biases from data collection. Firstly, the states designated as high capacity did undergo two rounds of National Deworming Day, opposed to one, so intuitively we would expect larger results. Although this does not explain why no significant results were found for states that only took part in 2016's Deworming Day, which reached many more children than in 2015. Secondly, the second round of deworming is documented to have included a small amount of private schools, just under 10% of all treatments, again underestimating the true impact of the treatment in 2016. The possibility of imperfect timing of data collection is also important. The ASER survey is an ongoing yearly project, so data collected in the first two months of the survey will not capture the effect of National Deworming Day in 2016. Finally, it is important to consider the theoretical mechanism by which the treatment impacts the outcomes being measured. One might expect the benefits of the deworming treatment to be far from immediate, especially regarding the return to schooling and educational outcomes. Keeping this in mind, the timing of the data collection may obscure the magnitude of longer-term results from the 2016 treatment, in line with findings from Ozier (2018) and Kremer & Miguel et al (2019)

Although the above mentioned sources of structural bias may account for a meaningful proportion of the differential results, it is still true that the high capacity states were uniquely successful in administering a universal and successive National Deworming Day. They demonstrate the significant impact of universal outreach and broad-based national programs in reducing infectious health risks. Moreoever, the results for the high-capacity states further show the effectiveness of the treatment generally.

In Table 7 I present results for the triple-difference estimator using a logistic regression model. These results act as a robustness check on the Linear Probability Model. Importantly, the results are presented in terms of log-odds ratios, rather than probability as in previous tables. In this way, the results are not easily interpretable as with the linear probability model. Instead, these results demonstrate the robustness of the result to alternative specifications.

Table 7: Triple-Difference Logit Model

	5-7			11-16		
	Reading	Math	English	Reading	Math	English
Treat	-0.4739***	-0.6225***	-0.9524***	-0.3087***	-0.4352***	-0.5479***
	(0.0220)	(0.0224)	(0.0241)	(0.0257)	(0.0248)	(0.0259)
Post	-0.0616**	-0.0887***	-0.0735***	-0.2000***	-0.2782***	-0.2560***
	(0.0250)	(0.0256)	(0.0266)	(0.0294)	(0.0289)	(0.0297)
NDD2015	-0.0157	-0.4014***	0.1276	-0.2758***	-0.5587***	-0.2618***
	(0.0865)	(0.0889)	(0.0941)	(0.1004)	(0.0964)	(0.1010)
Treat*Post	-0.0693**	-0.0431	0.0342	-0.0047	0.0016	0.0322
	(0.0311)	(0.0313)	(0.0336)	(0.0350)	(0.0342)	(0.0351)
NDD2015*Post	0.0571	0.1170***	0.1103***	0.0892**	0.2149***	0.1564***
	(0.0369)	(0.0376)	(0.0394)	(0.0420)	(0.0410)	(0.0415)
NDD2015*Treat	0.1593***	0.0969***	0.0477	0.1829***	0.1214***	0.0341
	(0.0309)	(0.0314)	(0.0337)	(0.0351)	(0.0337)	(0.0346)
Treat*Post*NDD2015	0.1128***	0.0754*	0.0348	0.0789	0.1025**	0.0869*
	(0.0432)	(0.0436)	(0.0463)	(0.0480)	(0.0464)	(0.0474)
β_0	0.6526***	1.2189***	1.3203***	1.5689***	1.3275***	1.6103***
	(0.0423)	(0.0443)	(0.0462)	(0.0516)	(0.0496)	(0.0513)
Number of Observations	439602	439602	439602	425745	425745	425745

Note: Coefficients are expressed in terms of log-odds. Controls include state, presence of clinics and Anganwadis, and household toilet access

7.4 Ordered Response Categories

To conclude empirical results of the paper, I present findings from the Generalized Ordered Logit Model, which estimates coefficients for the log odds of falling into numerically ordered response categories. I present these results for both the double-difference and triple-difference specification (Tables 8 & 9). The results are presented in terms of odds-ratios, which is simply the exponentiated form of log-odds-ratios.

The results from Table 8 and 9 offer a more complete accounting of the values of the test score data, and again show significant effects of the treatment on the cognitive abilities of treated children. Due to a lack of prior research on the use of double and triple-differences coefficients in ordered logistic regression models, I refrain from interpreting the results, but note the significance, namely that the treatment is found to primarily expected to significantly increase the probability of children higher test scores. The level at which results are significant however, varies by child and test type. The GOLOGIT model demonstrates significant positive effects for children aged 11-16 in all test score types, across both specifications. Interestingly, negative and significant effects are found for children 5-7's reading scores in low capacity-states, but significant and positive effects are found for the same group in high capacity states. We also see significant positive effects for Math and English scores in high capacity states. These findings demonstrate an alternative estimation strategy, and further evidence of the positive and significant effects of India's National Deworming Day on rural children's educational outcomes.

Table 8: Generalized Ordered Logit Model: Double-Difference

	5-7			11-16		
	Reading	Math	English	Reading	Math	English
OR: Score > 1	0.0308	0.0271	0.0367	0.1942***	0.2212**	0.0675
	(0.0296)	(0.0315)	(0.0291)	(0.0715)	(0.0892)	(0.0582)
OR: Score > 2	0.0233	0.0275	0.0713***	0.0922**	0.1160***	0.1222***
	(0.0228)	(0.0233)	(0.0255)	(0.0416)	(0.0423)	(0.0400)
OR: Score > 3	0.0100	0.0326	0.0928***	0.0013	0.0594**	0.0855***
	(0.0230)	(0.0251)	(0.0257)	(0.0334)	(0.0265)	(0.0287)
OR: Score > 4	-0.0035	-0.0145	0.0524	0.0101	0.0638**	0.0318
	(0.0256)	(0.0350)	(0.0347)	(0.0276)	(0.0248)	(0.0258)
Number of Observations	376489	376127	374701	350601	350238	349746

Table 9: Generalized Ordered Logit Model: Triple-Difference

	5-7			11-16		
	Reading	Math	English	Reading	Math	English
OR: Score > 1 (1)	-0.0232	0.0403	-0.0061	0.0989	0.2067*	-0.0135
	(0.0416)	(0.0436)	(0.0413)	(0.0987)	(0.1222)	(0.0809)
OR: Score > 1 (2)	0.0630	-0.0208	0.0802	0.1354	0.0678	0.2368**
	(0.0591)	(0.0632)	(0.0589)	(0.1432)	(0.1803)	(0.1167)
OR: Score > 2 (1)	-0.0573*	-0.0292	0.0529	0.0323	0.0625	0.1329**
	(0.0329)	(0.0334)	(0.0370)	(0.0562)	(0.0580)	(0.0570)
OR: Score > 2 (2)	0.1154**	0.0650	0.0205	0.0553	0.0937	-0.0522
	(0.0462)	(0.0471)	(0.0519)	(0.0845)	(0.0870)	(0.0809)
OR: Score > 3 (1)	-0.0709**	-0.0541	0.0370	-0.0648	-0.0210	0.0245
	(0.0342)	(0.0379)	(0.0375)	(0.0470)	(0.0395)	(0.0427)
OR: Score > 3 (2)	0.1447***	0.1323***	0.0989*	0.1228*	0.0769	0.0615
	(0.0472)	(0.0512)	(0.0520)	(0.0673)	(0.0535)	(0.0578)
OR: Score > 4 (1)	-0.0961**	-0.0219	0.0428	-0.0611	-0.0238	0.0027
	(0.0383)	(0.0530)	(0.0487)	(0.0404)	(0.0372)	(0.0386)
OR: Score > 4 (2)	0.1633***	0.0020	0.0522	0.1227**	0.1136**	0.0125
	(0.0526)	(0.0722)	(0.0698)	(0.0553)	(0.0501)	(0.0520)
Number of Observations	376489	376127	374701	350601	350238	349746

(1) refers to low-capacity states, (2) refers to high-capacity states

7.5 Analysis of Age

At this point in the paper I refer the reader to Figures 4-7 once again to display important descriptive statistics to contextualize the heterogeneity found in the results. Figure 7 displays that the attainment of the level-4 benchmark score is low (below 50%) amongst all younger-children (younger than 11) at baseline. Children 5-7 in particular, have an attainment rate clearly below 15 percent in all 3 testing categories. With this information in mind, it is important to note that a conceivable mechanism of the treatment is to assist students in meeting benchmarks that are reasonable for their age and level of learning. The lack of significant results for younger-aged children also represents an important theoretical concept. The theoretical mechanism of the treatment is to alleviate children from health burdens, restoring lost capabilities in educational outcomes. Then, we would expect to see less significant results from younger children who were not likely to attain such high results in the first place. The use of differential benchmarks by age is designed to adjust for this dynamic, yet it is unlikely to fully ameliorate it. This interpretation of the results then suggests that the deworming treatment is critical to restoring cognitive abilities of children that otherwise lost due to the debilitating or distracting nature of soil-transmitted helminths. This fits well with the motivating theory of the paper, that infection by STHs leads to a significant loss of human capital that would counterfactually be present.

From an empirical standpoint, this analysis may also bring the reader to another important conclusion. A public health treatment cannot be expected to raise educational outcomes of a group higher than a level obtained by members of that same group who have no need for the treatment. To further illustrate this point I refer again to the histograms of test scores by age group for the reader (Figures 4-6). Viewing the results in the context of these distributions lends additional support to this theoretical mechanism.

8 Discussion and Conclusion

The results shown thus far indicate that India's National Deworming Day had immediate and tangible effects on the educational attainment of its youth population. Strikingly, the significant positive impacts were found primarily in states designated as having the capacity to effectively carry out such a project. Before confirming the inference of the results however, it is important to discuss the challenges and shortcomings of the study.

Firstly, the extreme scope of this study all but precludes the possibility of rigorously verifying the validity of the parallel trends assumption. Although the analysis of trends above is compelling, given the size of study encompassing all of India, it is extremely difficult to verify that no additional programs or trends were endogenous to treatment status during the treatment period. Thus, the estimated treatment effects could conceivably be inflated or deflated if any additional unaccounted shocks affected the treatment and control groups unevenly.

Another critical consideration stems from limitations with regards to data availability. The nature of the treatment variable somewhat complicates the interpretation of the results. Because treatment is defined by school enrollment, treatment effects on school attendance can not be reliably estimated, meaning that any changes in test scores can not be disentangled changes in school attendance, which are clearly not independent. Furthermore, the survey only provides data on school enrollment, not school attendance. Fortunately, the study is designed to internalize this fact. Since the study attempts to estimate the impact of National Deworming Day on national educational outcomes, indirect improvements in educational outcomes that are the result of improvements in school attendance are just as important as direct effects on educational outcomes. Although more sophisticated analyses could be carried out if school attendance could be measured accurately, the lack of such data does not harm the inference of the reported results.

The findings align well with the theoretical motivations for the paper. For average rural Indian children aged 5-10, I estimated that the treatment increased the probability identifying lowercase English letters by just over 1.5 percentage points. For children aged 11-16, the study estimates that India's National Deworming Day increased the probability of successfully reading a standard level-1 text by 1.4 percentage points, increased the probability of successfully doing two-digit subtraction by 2.4 percentage points, and increased the probability of correctly reading simple English words by 2.4 percentage points as well.

The effects were also shown to be highly concentrated in states designated by the Ministry of Health and Family Welfare as being well-prepared to conduct a mass, universal deworming program. I found larger and more significant treatment effects for these states specifically, and null effects for all other states. These results offer meaningful support to the theoretical claim that high levels of state capacity better prepare governments for launching ambitious programs, specifically when attempting to curb the spread of an infectious disease like STHs. Although this paper offers no measurement of "state capacity", nor any continuous relationship between state capacity and the outcomes of interest, the paper acts as a valuable contribution to the literature regarding the role of state capacity in public health programs, disease eradication, and national development.

Importantly, I see large effects by age group on the probability of obtaining an outcome that has been attained by significant portion of that age group. For example, a very small proportion of children aged 5-10 had attained the level 4 benchmark at baseline, and thus the estimated treatment effects for children of that age were not statistically different from zero. By comparison however, treatment effects for older age groups, who had higher levels of baseline attainment of the benchmark of 4, were significantly higher in magnitude and statistical significance. This aligns well with the model of the S-shaped poverty trap.

These findings suggest that the deworming treatment plays a minimal role in directly raising cognitive abilities. Rather, the deworming treatment is pivotal to improving educational outcomes due to its mitigating effects on infection by soil-transmitted helminths. Then, the primary impact of the treatment is the restoration of human capital that is expended on or damaged by the infection. Removing STH infections is almost certain to send more children to school, where they will learn more, ceteris paribus, than if they had not attended. For children already attending school, the discomfort of STHs may be distracting, or the loss of nutritional absorption may impair their human capital. In this way, removing the infection by way of the treatment is an important restoration of children's human capital.

The results of the study aligns very well with models of externalities associated with treatment. Given the estimation strategy and the size of the study, the estimation of such significant treatment effects lends support to the theory that externalities comprise a large portion of the benefits from treatment. Additionally, the findings lend strong support to the notion of an S-shaped or nutrition-based poverty trap. The heterogeneity in treatment effects by age indicate that for many children, the deworming treatment allows them to "catch-up" to their better performing peers, implying that their own productivity, or human capital, is impaired by STH infection. The deworming treatment then, acts as a valuable investment in human capital, which would be expected to translate to long-term increases in lifetime earnings. Due to the short-term nature of this study, that hypothesis is not able to be tested, but the theoretical motivations follow from the study's short-term findings.

Additionally, the results offer a valuable contribution to existing literature on government efficacy and state capacity, particularly regarding public health interventions and the impact of universal programs. The infectious nature of STHs points to the importance of broad universal programs like National Deworming Day to the minimize transmission and health risks associated with the infection. Although the designation of state capacity was a simplistic binary one, it was made by experts in the Indian Ministry of Health and Family Welfare. Importantly, we saw that public agencies have the ability to effectively measure and predict levels of state capacity, and that states with this capacity performed significantly better in administering the treatment. These results add further contributions to literature on State Capacity, providing limited evidence that some governments are able to successfully designate meaningful levels of State Capacity that can have significant impacts on the success and administration of public programs.

Finally, the findings offer further, and significant support to the notion of private under-consumption of the deworming treatment, and the value of public provision of such goods. The deworming treatment has been widely publicly available and affordable for years, and shown to have significant positive impacts on health, human capital, and wages since as early as 2004, yet it continues to be chronically under-consumed. The finding that the spontaneous public treatment in 2015 and 2016 had significant impacts on children's human capital is significant evidence of private under-consumption of the deworming treatment.

References

[1] **Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson**. "State Capacity and Economic Development: A Network Approach." *American Economic Review*, 105 (8): 2364-2409.

[2] **Baird, S., Hicks, J. H., Kremer, M., Miguel, E.**. "Worms at Work: Long-run Impacts of a Child Health Investment." *The Quarterly Journal of Economics*. November 2016, 131 (4), 1637–1680.

[3] **Ben Jann**. "Plotting regression coefficients and other estimates." *Stata Journal*. StataCorp LP. Vol. 14(4). pages 708-737. December 2014.

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

[5] **Card, David, and Alan B. Krueger.** Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 1994, 84, Issue 4, p. 772-93.

[6] **Dasgupta, Partha, and Debraj Ray.** "Inequality as a Determinant of Malnutrition and Unemployment: Theory." *The Economic Journal* 1986, 96, no. 384 (1986): 1011-034. Accessed February 14, 2021. doi:10.2307/2233171.

[7] **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.

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

[9] **Habyarimana, J., Jack, W.** "Results of a large-scale randomized behavior change intervention on road safety in Kenya." *Proceedings of the National Academy of Sciences*, 2015, 112(34), E4661 LP-E4670. https://doi.org/10.1073/pnas.1422009112

[10] **Jessica Cohen, Pascaline Dupas**, "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." *The Quarterly Journal of Economics*, Volume 125, Issue 1, February 2010, Pages 1–45, https://doi.org/10.1162/qjec.2010.125.1.1

[11] **Joachim Weimann, Jeannette Brosig-Koch, Timo Heinrich, Heike Hennig-Schmidt, Claudia Keser**. "Public good provision by large groups – the logic of collective action revisited." *European Economic Review*, Volume 118, 2019, Pages 348-363, ISSN 0014-2921, https://doi.org/10.1016/j.euroecorev.2019.05.019.

[12] **Kirwan, Patrick, Andrew Jackson, Samuel Asaolu, S´ıle Molloy, Titilayo Abiona, Marian Bruce, Lisa Ranford-Cartwright, Sandra O'Neill, and Celia Holland.** "Impact of repeated four-monthly anthelmintic treatment on Plasmodium infec- tion in preschool children: A double-blind placebo-controlled randomized trial." *BMC Infectious Diseases* 2010 (September): 277

[13] **Kristoffer Bjärkefur, Luiza Cardoso de Andrade, Benjamin Daniels, Mrijan Rimal**, "IETOOLKIT: STATA toolkit of commands for impact evaluations." *DIME Analytics*. revised 08 Nov 2019.

[14] **Larocque, R., M. Casapia, E. Gotuzzo, J.D. MacLean, J.C. Soto, Rahme E., and T.W. Gyorkos.** "A double-blind randomized controlled trial of antenatal mebendazole to reduce low birthweight in a hookworm-endemic area of Peru." *Tropical Medicine International Health* 2006, 11 (10): 1485–1495.

[15] **Mazumdar, Dipak**. "The Marginal Productivity Theory of Wages and Disguised Unemployment." *Review of Economic Studies*, 1959, Oxford University Press, vol. 26(3), pages 190-197.

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

[17] **Mincer, Jacob.** "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy* 66, no. 4 (1958): 281-302. Accessed February 14, 2021. http://www.jstor.org/stable/1827422.

[18] **Ozier, Owen.**, "Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming," *American Economic Journal: Applied Economics.* July 2017, 10 (3), 235-62.

[19] **Pratham**. Annual Status of Education Report 2011–2016. Mumbai: *Pratham Resource Centre*. Google Scholar

[20] **Pullan, R.L., Smith, J.L., Jasrasaria, R. et al.** "Global numbers of infection and disease burden of soil transmitted helminth infections in 2010." *Parasites Vectors 7*, 37 2014. https://doi.org/10.1186/1756-3305-7-37

[21] **Ravallion, Martin** "The Idea of Antipoverty Policy." NBER Working Papers 19210, *National Bureau of Economic Research*, 2013, Inc.

[22] **Sachs, Jeffrey.** *The End of Poverty: Economic Possibilities for Our Time.* New York: Penguin Books, 2006. Print.

[23] **Singh N.**. "Decentralization and public delivery of health care services in India." *Health Affairs* (Millwood). 2008 Jul-Aug ; 27(4) : 9 91-1001. doi: 10.1377/hlthaff.27.4.991. PMID: 18607032.

[24] **Stephenson LS, Latham MC, Adams EJ, Kinoti SN, Pertet A.** "Physical fitness, growth and appetite of Kenyan school boys with hookworm, Trichuris trichiura and Ascaris lumbricoides infections are improved four months after a single dose of albendazole." *J Nutr.* 1993 Jun;123(6):1036-46. doi: 10.1093/jn/123.6.1036. PMID: 8505663.

[25] **Stoltzfus, R.J., H.M. Chwaya, Tielsch J.M., Schulze K.J., Albonico M., and Savi- oli L.** . "Epidemiology of iron deficiency anemia in Zanzibari schoolchildren: the importance of hookworms" [in English (US)]. *American Journal of Clinical Nutrition* 1997, 65, no. 1 (January): 153–159.

[26] **Wammes, Linda J., Firdaus Hamid, Aprilianto Eddy Wiria, Linda May, Maria M. M. Kaisar, Margaretta A. Prasetyani-Gieseler, Yenny Djuardi, et al.** "Deworming improves immune responsiveness." *Proceedings of the National Academy of Sciences* 2016, 113 (44): 12526–12531.

[27] **Williams, Richard**. "Understanding and interpreting generalized ordered logit models." *The Journal of Mathematical Sociology*, 2016, 40:1, 7-20, DOI: 10.1080/0022250X.2015.1112384.

[28] **Williams, Richard**. "Generalized Ordered Logit/ Partial Proportional Odds Models for Ordinal Dependent Variables." *The Stata Journal*, 2006, 6(1): 58-82.

[29] **World Bank Group**. "What Is State Capacity?" Policy Research working paper; no. WPS 8734 Washington, D.C. : *World Bank Group.* http://documents.worldbank.org/curated/en/336421549909150048/What-Is-State-Capacity