NBER WORKING PAPER SERIES

META-ANALYSIS AND PUBLIC POLICY: RECONCILING THE EVIDENCE ON DEWORMING

Kevin Croke
Joan Hamory Hicks
Eric Hsu
Michael Kremer
Ricardo Maertens
Edward Miguel
Witold Więcek

Working Paper 22382
http://www.nber.org/papers/w22382

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016, Revised March 2022

We are grateful to the Campbell Collaboration deworming team, especially Vivian Welch, for generously sharing data and information on their work. We thank Harold Alderman, Shally Awasthi, Don Bundy, Jaime Carmona-Fonseca, Serene Joseph, Chengfang Liu, Scott Rozelle, and Walter Willett for help interpreting their studies, searching for original data, or supplying us with additional results. We thank Egor Abramov, Arthur Baker, Alexandre Simoes Gomes, Yue (Luna) Huang, Kamran Jamil, Spencer Ma, Cole Scanlon, Wen Wang, and Kevin Xie for research assistance. We thank Amrita Ahuja, Harold Alderman, Ben Arnold, Sarah Baird, Matthew Basilico, Jade Benjamin-Chung, Jack Colford, Peter Hotez, Rachael Meager, Antonio Montresor, Owen Ozier, Caio Piza, Dina Pomeranz, Gil Shapira, and many seminar and conference participants, for helpful comments. We thank the Dioraphte Foundation for financial support. All errors are our own. *Ricardo Maertens currently works at Amazon. He contributed to this research paper prior to joining Amazon. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Meta-Analysis and Public Policy: Reconciling the Evidence on Deworming
Kevin Croke, Joan Hamory Hicks, Eric Hsu, Michael Kremer, Ricardo Maertens, Edward Miguel, and Witold Więcek
NBER Working Paper No. 22382
July 2016, Revised March 2022
JEL No. C49,I15,I18,O15

ABSTRACT

The WHO recommends mass drug administration (MDA) in areas with >20% prevalence. Recent Cochrane meta-analyses endorse treatment of infected individuals but recommend against MDA. A theory-agnostic meta-analysis of the effect of multiple-dose MDA rejects the hypothesis of a common zero effect on child weight, mid-upper arm circumference, and height, and estimates significant average impacts. Estimates of implied treatment effect on infected children with MDA are not significantly different than those found for test and treat trials. These results suggest that MDA is a cost-effective intervention, particularly in the settings recommended by the WHO.

Kevin Croke
World Bank
1818 H Street, NW
Washington DC 20433
kcroke@worldbank.org

Joan Hamory Hicks
Department of Economics
University of Oklahoma
308 Cate Center Drive
Norman, OK 73072
jhamory@ou.edu

Eric Hsu
Department of Economics
Evans Hall, #3880
Berkeley, CA 94720-3880
eric.hsu@berkeley.edu

Michael Kremer
University of Chicago
Kenneth C. Griffin Department of Economics
1126 E. 59th St.
Chicago, IL 60637
and NBER
kremermr@uchicago.edu

Ricardo Maertens
Amazon
515 Westlake Ave N
Seattle, WA 98109
ricardo.maertens@gmail.com

Edward Miguel
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
emiguel@econ.berkeley.edu

Witold Więcek
Development Innovation Lab
University of Chicago
witold.wiecek@gmail.com

A data appendix is available at https://data.nber.org/data-appendix/w22382/

1 Introduction

Evidence from randomized controlled trials (RCTs) is accumulating in economics, and meta-analysis is increasingly being used to help interpret this growing body of evidence, including for public policy (Vivalt, 2015, 2020; Meager, 2018, 2020). While such studies can be very helpful in informing policy design, they can also be conducted or interpreted in ways which do not directly inform, and may even mislead, policy decisions. In this article, we report on several approaches to meta-analyses of the impact of mass treatment of intestinal worms. We conduct meta-analyses which directly address policy decision problems, while also addressing some of the limitations of previous studies. The results help resolve the apparent paradox, seen in recent work on mass drug administration (MDA), that trials of treatment of infected children find large positive impacts, while meta-analysis suggests that mass treatment of populations that include infected children does not have beneficial effects.

The World Health Organization (WHO) has long recommended MDA for intestinal worms among children in areas with more than 20% prevalence of intestinal worms (1 annual dose), or more than 50% prevalence (2 annual doses). However, recent meta-analyses have cast doubt on this recommendation. Taylor-Robinson et al. (2015) estimate that single dose treatment for children known to be infected leads to statistically significant gains across various nutritional outcomes, and express support for treating these children (p. 30). However, they argue that there is “substantial evidence” that mass drug administration (MDA) has no impact on child outcomes (p. 3) and recommend against its implementation in poor regions (p. 30). This creates an apparent paradox: if infected individuals benefit then one would expect a smaller, but still positive, average effect of MDA in endemic populations. The paradox remains in the 2019 update of the review, where the authors argue that it is “obvious” (p. 29) that children known to be infected with worms should receive treatment, but reaffirm their recommendation against mass treatment in endemic populations (Taylor-Robinson et al., 2019).

In this paper, we first follow Taylor-Robinson et al. (2015, 2019) and conduct a theory-agnostic meta-analysis of trials of multiple-dose MDA with outcomes measured at longest follow-up. We limit the analysis to trials that report effects on children's weight, mid-upper arm circumference (MUAC), height, or hemoglobin (Hb). By following the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green, 2011) we are able to strengthen the analysis in several ways. First, we include studies that were identified by Taylor-Robinson et al. (2015, 2019) but were excluded from their meta-analysis (for instance, because standard errors were not directly reported in the study but could be calculated from other reported statistics). Second, we extract point estimates and standard errors of the impact of deworming MDA using the most precise estimators available (e.g., ANCOVA, difference-in-differences). As a consequence, the statistical analysis is better powered to detect nutritional gains from deworming than previous meta-analyses (Taylor-Robinson et al. 2015, 2019; Welch et al. 2016).

The results help resolve the paradox in the deworming literature. The hypothesis that deworming MDA has a zero effect in all settings can be rejected for weight (p-val<0.001), MUAC (p-val<0.001), and height (p-val<0.05). We can also reject the hypothesis that the effects of MDA on the four outcomes analyzed are jointly zero (p-val<0.001). Focusing on settings where the WHO currently recommends MDA (>20% and >50% prevalence) and using a random effects model, we estimate significant average increases in child weight of 0.154 kg; in MUAC of 0.198 cm; and in height of 0.087 cm, for over 20% prevalence settings, and 0.172 kg, 0.198 cm, and 0.095 cm, for over 50% prevalence settings. There is a positive but insignificant effect on Hb of 0.069 g/dl (p-val=0.073), for over 20% prevalence settings, and 0.044 g/dl (p-val=0.589), for over

50% prevalence settings. We note that low Hb and anemia are linked to hookworm infection; however only two settings in the Hb sample have hookworm prevalence of over 20% (Carmona-Fonseca & Correa-Botero, 2015; Ostwald et al., 1984), limiting the statistical power of this test.

Damage from worms, i.e., morbidity, is generally agreed to depend on infection intensity, which is highly skewed and strongly positively correlated with disease prevalence: in low prevalence settings, relatively few people have severe infections, while many more do in high prevalence populations (World Health Organization, 2017). Although we do not directly include infection intensity in the analysis of treatment effect due to the lack of conclusive evidence regarding the nature of this relationship, we are at least able to explore treatment effects by infection prevalence. We estimate the implied effect of MDA on infected children and compare it with the mean effect on screened children in test and treat trials. We are not able to reject the hypothesis that these two effects are equal, helping to reconcile the results from these two groups of studies.

We next argue that when interpreting a body of evidence, it is important to incorporate a decision theory perspective. We show that the estimated health gains per dollar spent from deworming MDA (in settings with over 20% worm prevalence) are over 23 times as large for weight, 50 times as large for MUAC, and 13 times as large for height, relative to those of school feeding, another widely implemented intervention that targets similar outcomes in similar populations (Kristjansson et al., 2007; Kristjansson et al., 2016). For settings with over 50% worm prevalence, these values are 25 for weight, 50 for MUAC, and 13.6 for height. Therefore, it makes sense to condition the decision of whether to implement MDA, and the choice of the number of doses, on the characteristics (prevalence) of the setting under analysis. These calculations echo a recent epidemiological study that finds deworming to be highly cost-effective (Lo et al., 2016). A

Bayesian analysis suggests that policymakers would need extremely confident priors that MDA has no effect in order to believe that it is not more cost effective than school feeding. Furthermore, we cannot reject the hypothesis (at conventional levels) that deworming MDA has non-negative effects everywhere, for any of the outcomes examined in this paper. This is relevant for policymakers who would be reluctant to use MDA if there were evidence that it could have negative average effects in some settings.

This paper is organized as follows. Section 2 provides background information on soil-transmitted helminths, mass drug administration, and the economics literature on deworming. Section 3 discusses the sample, the criteria for study inclusion, the procedure used to identify studies, and the general principles guiding data extraction. Section 4 describes the statistical methods we use for hypothesis testing, estimation, cost-effectiveness analysis, and prediction. Section 5 presents the results, while Section 6 revisits other meta-analyses by Taylor-Robinson et al. (2015, 2019) and Welch et al. (2016). Section 7 concludes.

2 Background

Soil transmitted helminths (STH; including hookworm, whipworm, and roundworm) are among the most widespread infectious diseases, affecting 1 in 4 people in endemic countries (Pullan et al., 2014). STH are spread via eggs deposited in the local environment through feces. School-aged children are especially vulnerable to infections and play an important role in spreading them locally (Hotez et al., 2006). These infections affect child health and nutrition through impaired nutritional intake, reduced nutrient absorption, intestinal damage, dysentery, blood loss, and combinations of these pathways, depending on the worm species (WHO, 2017). For example, hookworms create lesions in their host’s intestinal mucosa, leading to blood loss and low hemoglobin levels.

Hookworm infection is among the leading five causes of anemia globally (Kassebaum et al., 2014). Morbidity is generally agreed to depend on infection intensity as measured by the number of worms in the infected host (as measured by the number of worm eggs per gram of feces), rather than a simple binary infection indicator. Infection intensity is highly skewed and strongly positively correlated with disease prevalence: severe infections are much more common in high prevalence settings (Anderson, Truscott and Hollingsworth, 2014).

The most common drugs used to treat STH are albendazole and mebendazole (WHO, 2017), which have been found to be “extremely well tolerated” by infected and non-infected individuals (Albonico et al., 2008). Side effects from treatment are very infrequent (about 1%), minor (e.g., nausea, rashes, migration of worms through the mouth), mainly related to the elimination of heavy worm loads, and typically disappear within 48 hours (Joseph et al., 2016; Albonico et al., 2008; WHO, 2002).

There is agreement that children known to be infected should be treated; indeed, this is the standard of medical care (Horton, 2000; Keiser and Utzinger, 2008; Perez del Villar et al., 2012). Because deworming treatments are inexpensive and safe, but diagnosing infection is comparatively expensive and often logistically difficult, the WHO recommends annual mass treatment in areas where worm prevalence is above 20% and biannual treatments where prevalence is greater than 50%. Screening for worm infections requires testing stool samples, which in turn necessitates skilled staff, laboratory facilities, and re-contacting infected individuals for treatment. Moreover, standard testing methods have an estimated sensitivity between 52% and 91% (Barda et al., 2013; Assefa et al., 2014), suggesting that many infections would go undetected even with screening. Further, the cost of screening for worm infections is 4 to 10 times that of treatment (Taylor-Robinson et al., 2015, p. 7).

Subsequent to the WHO recommendation a social science literature emerged measuring the long-term educational and economic impacts of mass deworming, suggesting that the benefits of MDA far exceed the costs (Ahuja et al., 2015; Baird et al., 2016; Bleakley, 2007; Hamory et al., 2021). Three studies in moderate to high prevalence settings — in Kenya and the historical southern United States — find substantial long-run impacts of deworming on educational outcomes (Ozier, 2018; Bleakley, 2007; Baird et al., 2016). Several of these studies also report economic outcomes and find positive effects.[1] In addition, multiple organizations have ranked MDA as highly cost-effective.[2]

3 Sample and data extraction procedures

This section describes the trial inclusion criteria, the search procedure for identifying studies, and the procedures for extracting data from included trials.

We restrict the analysis to randomized controlled trials of MDA in which multiple doses of deworming treatment were administered and include treatment effect estimates from the longest

[1] Ozier (2018) finds that infants who lived in Kenyan communities where older school-age children were dewormed show large cognitive gains ten years later. Bleakley (2007) finds that deworming campaigns in the U.S. South in the early 1900s increased school enrollment and attendance, and increased literacy and income for adults who were treated as children; Roodman (2017a) and Bleakley (2018) discuss the robustness of these results to the inclusion of additional census data. Baird et al. (2016) estimate that a decade after treatment, males who participated in mass deworming in Kenya worked 17% more hours per week and had higher living standards. Females were approximately one-quarter more likely to have passed the primary-school leaving exam and attended secondary school. The estimated value of benefits, in terms of the net present value of future earnings net of increased schooling costs, exceeds the cost by more than one hundred-fold. Hamory et al. (2021) study the same population in Kenya and estimate that fifteen to twenty years after treatment, treated individuals experienced a 14% gain in consumption expenditures and a 13% gain in hourly earnings, implying that the deworming intervention had an annualized social rate of return of at least 37%. Another study (Croke and Atun, 2019) found that a mass deworming campaign in Uganda did not have significant average educational effects among the entire population.

[2] E.g., the Copenhagen Consensus (Hall and Horton, 2008), the Disease Control Priorities Project (DCPP; Hotez et al., 2006), Givewell (2014), Abdul Latif Jameel Poverty Action Lab (J-PAL Policy Bulletin, 2012), and the World Bank (1993).

follow-up reported.[3,4] The main analysis includes trials with any of the following child nutrition indicators as outcomes: weight, MUAC, height, or hemoglobin. We focus on these outcomes because for each, we were able to identify at least three studies examining the effects of multiple dose deworming. Only RCTs for which a causal intention-to-treat estimate can be obtained are included. Therefore, we require that the study report outcomes for the population assigned to treatment and comparison groups, independent of whether they received treatment or not.

When estimating the mean effect of MDA on child nutrition indicators, we report results both in the set of trials that take place in settings where the WHO recommends deworming (i.e., those where the baseline prevalence of hookworm, whipworm, or roundworm is over 20%, which is the threshold for annual MDA), and in the full sample. When examining the evidence on deworming infected children, we pool evidence from deworming MDA trials with that of deworming trials of children who were screened for infection ("test-and-treat" trials).

3.1 Search procedure

We start with the sample of studies identified by Taylor-Robinson et al. (2015), for their analyses of the impact of multiple-dose deworming treatment of "all children living in an endemic area" (i.e., mass drug administration, or MDA) at longest follow-up on children's weight, MUAC,

[3] This corresponds to what Taylor-Robinson et al. (2015, p. 4) term their "main comparison." In a subsequent meta-analysis, Taylor-Robinson et al. (2019) broaden the main category of analysis to allow for the inclusion of multiple-dose trials that screened children for infection. However, since Taylor-Robinson et al. (2019) identified no such trial, the main category of analysis remained de facto the same in the updated review. Taylor-Robinson et al. (2015) identified only one multiple dose "test-and-treat" trial in each outcome category we examine, namely, Stephenson et al. (1993). This trial is, in fact, an MDA trial and is classified as such in the 2019 update

[4] In other analyses, Taylor-Robinson et al. (2015, 2019) examine the effect of deworming after the first dose of treatment by combining data from multiple-dose MDA trials where effects are reported after the first dose with MDA trials of single-dose and, in the updated review, with single-dose trials that screened children for infection. We exclude single-dose MDA trials from the analysis as their length of follow-up are typically too short to allow for nutritional gains to emerge. For example, Taylor-Robinson et al. (2015, 2019) include Hadju et al. (1996) and Palupi et al. (1997) which are single-dose MDA trials with follow-up periods of 7 and 9 weeks, respectively. The median length of follow-up for multiple-dose MDA trials is one year.

height, and hemoglobin. We supplement this sample with additional studies of multiple-dose MDA identified by Welch et al. (2016) that meet the trial inclusion criteria above, and we update the systematic search for trials by Taylor-Robinson et al. (2015) to identify studies published between April 14, 2015 (the Taylor-Robinson et al. search date) and June 29, 2018.[5, 6] We also identify "test-and-treat" trials, following the study search and data extraction by Taylor Robinson et al. (2015).

3.2 Data extraction and choice of estimator

Data extraction follows six principles derived from the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green, 2011), which can help improve statistical power in meta-analysis and which we present below.

i. If treatment effects are presented without standard errors, standard errors are calculated using other presented data (e.g., t-statistics, p-values, or 95% confidence intervals), following the formulas provided in *The Cochrane Handbook* where possible (Higgins and Green 2011, section 7.7.3.3).

ii. If results are reported in figures rather than in the text or in a table, Web Plot Digitizer software (Rohatgi, 2015) is used to extract numerical estimates from the figures.

iii. If key information on treatment impacts is missing from a paper (and cannot be derived from what is presented), original microdata (where available) is used to obtain estimates.[7]

[5] See Appendix A.9 for details.

[6] Taylor-Robinson et al. (2019) acknowledge that an earlier working paper version of this study (Croke et al., 2016) and the Campbell review (Welch et al., 2016) made them aware of four studies they had not included in Taylor-Robinson et al. (2015) and included them in their updated review. They do not identify any additional trials which meet our trial inclusion criteria.

[7] We also obtain information from trial authors in several cases, through either direct communication or thanks to the generosity of the Campbell Collaboration research team.

iv. When possible, treatment effect estimates are extracted based on an Analysis of Covariance (ANCOVA) model. *The Cochrane Handbook* states that since ANCOVA estimates "give the most precise and least biased estimates of treatment effects they should be included in the analysis when they are available" (Higgins and Green (2011), section 9.4.5.2). When it is not possible to extract ANCOVA estimates, but it is possible to extract estimates based on changes from baseline to endline, this "difference-in-differences" or "changes" estimator is used. This estimator is typically more precise than the estimator based only on comparison of endline differences.[8]

v. In case of textual contradictions about key parameter values in a trial (for example, a study text that reports significant effects versus reported test statistics that imply non-significant results), we first try to obtain the original microdata to perform the estimation ourselves. Where this is not possible, we assess which statistics were the primary focus of reporting in the text. In such cases, we also contact the original authors for clarification.

vi. Where studies report multiple treatment estimates, we follow the standard in Taylor-Robinson et al. (2015) and the medical literature of favoring unadjusted estimates. If studies do not report unadjusted estimates, but they do report treatment effects adjusted with standard covariates or baseline values (such as age and sex), these estimates are included in the analysis.[9]

[8] When outcomes are highly autocorrelated over time, estimators that take into account baseline information remain unbiased, while typically improving precision, and thus are preferable under standard statistical criteria, such as the goal of minimizing mean squared error (McKenzie, 2012). Following *The Cochrane Handbook*, when baseline and endline means and measures of variance were present but variance of the changes are missing, the standard error for changes is calculated using a correlation coefficient for the value between baseline and endline imputed from other studies (Higgins and Green 2011, section 16.1.3.2).

[9] Since expected changes in nutrition vary with age, including age as a covariate should generally improve precision, and should not induce bias (McKenzie, 2012).

Appendix Table A.1 presents a summary of the differences between the sample used in this paper and that of Taylor-Robinson et al. (2019). Taylor-Robinson et al. (2019) excluded 5 studies that we include in this meta-analysis, and include one study that we exclude. This paper's sample further differs from that of Taylor-Robinson et al. (2019) in that we extract different estimates from another eight studies, following the data extraction principles i-vi outlined above. These differences are discussed further in Section 6, and more details are presented in Appendix A. For each of the included studies we defined a prevalence variable, defined as maximum of prevalences over all worms reported in that study. In three studies this quantity had to be imputed. Details are provided in Appendix A.8.

Table 1 presents data on all outcomes and prevalence values used for statistical modeling. The sample of the child nutrition effects of MDA for deworming includes 27 estimates (22 trials) for weight, 7 estimates (6 trials) for MUAC, 22 estimates (17 trials) for height, and 13 estimates (9 trials) for hemoglobin.[10] Dispersion of mean treatment effects is large across all outcomes (weight: from -0.5 to 0.9 kg; height: from -1.2 to 1.4 cm; MUAC: from -0.4 to 0.8 cm; Hb: from -0.1 to 0.3 g/dl) as well as prevalences (from 3 to 95% among MDA studies: 6 MDA studies have less than 20% prevalence, 6 studies have between 20% and 50%, 19 studies have more than 50%).

4 Statistical models and methods for meta-analysis

Subsection 4.1 presents tests of the hypothesis that the effects of MDA on child nutrition are zero in all settings. Subsection 4.2 presents frequentist methods to estimate the mean effect of

[10] Because hemoglobin and anemia have been linked to hookworm infection (WHO, 2017) and because only two settings in the hemoglobin sample have hookworm prevalence of over 20% (Carmona-Fonseca & Correa-Botero, 2015; Ostwald et al., 1984), we do not expect to detect positive effects on hemoglobin. In addition, the median length of follow-up in the sample is 1 year. Because height is typically considered to reflect a person's cumulative nutritional status over time, longer treatment periods might be needed to detect effects on height.

deworming MDA assuming random effects and uses infection prevalence to compare the effects of MDA trials with the effects of test and treat trials on infected children. Subsection 4.3 links the evidence to a decision theory, focusing on cost-effectiveness, and outlines a framework to examine (i) whether MDA is cost-effective on average relative to a comparator intervention (school feeding), (ii) how pessimistic a policymaker would have to be for MDA not to be cost-effective, and (iii) whether one can reject the hypothesis that the effects of MDA are non-negative in all settings.

Throughout, we adopt the notational convention that for a given outcome, indexed by k ($k = 1, \dots, K$), the decision maker has access to point estimates ($\hat{\theta}_{j,k}$) and standard errors ($\hat{\sigma}_{j,k}$) of the effect of deworming MDA, coming from s_k settings, indexed by j ($j = 1, \dots, s_k$).

4.1 Testing the hypothesis of a zero effect across all settings and fixed-effect estimation

In light of the conclusion by Taylor-Robinson et al. (2015, p. 3) that there is “substantial evidence” of no impact of deworming MDA, we test the hypothesis that the true impact of multiple-dose deworming is zero in all settings ($\theta_{j,k} = 0,\ \forall j$). We first test this hypothesis against the alternative of a non-zero effect in at least one setting ($\theta_{j,k} \neq 0$ for some j), using the test statistic $\sum_{j=1}^{s_k} \hat{\theta}_{j,k}^2 / \hat{\sigma}_{j,k}^2$, which follows a $\chi_{s_k}^2$ distribution.

This test, however, has been shown to have relatively low power (Higgins et al., 2009). A higher-powered test of the null hypothesis of a common zero effect on outcome k can be conducted assuming a common effect of MDA across settings ($\theta_{j,k} = \theta_k,\ \forall j$) and conducting fixed-effect estimation of the common effect. One can then use the $Z_k = \hat{\theta}_k^{FE} / \hat{\sigma}_k^{FE}$ test statistic to test the hypothesis of a common zero effect ($\theta_k = 0$) against the alternative of a common non-zero effect

($\theta_k \neq 0$), where $\hat{\theta}_k^{FE}$ is the fixed-effect estimate, $\hat{\sigma}_k^{FE}$ is its standard error, and Z_k follows a standard normal distribution.[11]

In addition, the vector of fixed-effect estimators can be used to test the hypothesis that the common effect of mass deworming is zero across all outcomes. Denote by $\hat{\theta}^{FE}$ the K-dimensional vector of fixed-effect estimators and denote its variance-covariance matrix by Σ. The null hypothesis that $\theta = [0]_{K\times 1}$ can be tested using a Wald test, where the test statistic $\hat{\theta}^{FE'}\Sigma^{-1}\hat{\theta}^{FE}$ follows a χ_K^2 distribution (Wooldridge (2010), section 12.6.1). To implement this test, however, we need data on the covariances of the fixed-effect estimators. Because we do not have a strong prior about Σ, we assume that all pairs of fixed-effect estimators have the same correlation coefficient ($\rho = \frac{cov\left(\hat{\theta}_k^{FE},\hat{\theta}_{k'}^{FE}\right)}{\hat{\sigma}_k^{FE}\hat{\sigma}_{k'}^{FE}}$, $\forall k \neq k'$) and test the hypothesis that the common effect of mass deworming is zero across all outcomes, for various levels of the correlation coefficient. Under these assumptions, the correlation coefficient will be bounded by the following restriction on the covariances: (i) $-\hat{\sigma}_k^{FE}\hat{\sigma}_{k'}^{FE} \leq cov\left(\hat{\theta}_k^{FE},\hat{\theta}_{k'}^{FE}\right) \leq \hat{\sigma}_k^{FE}\hat{\sigma}_{k'}^{FE}$, $\forall k \neq k'$ and (ii) Σ needs to be positive semi-definite. Given that Σ is no longer positive semi-definite for very low and high values of the correlation, we bound the correlation coefficient by the minimum and maximum levels for which Σ is positive semi-definite.[12]

[11] Rice et al., (2018) notes that even if the assumption of a common effect does not hold, the Z statistic still allows for testing the hypothesis that the precision-weighted, average effect of mass deworming across settings is equal to zero.

[12] Andrews and Kasy (2019) estimate that negative estimates of the effect of MDA on weight are ten times less likely to be published than are positive ones, suggesting that estimates of deworming's impact based on published estimates may be upward biased. This issue is discussed in Appendix H.

4.2 Meta-analysis models

The mean effect of deworming MDA can be consistently estimated through random-effects estimation if studies are exchangeable, i.e., if the true effects of deworming can be assumed to be random draws from a common distribution of effects. While impacts may vary across settings as a function of covariates--and in fact the samples of trials generally exhibit heterogeneity in effect sizes, likely driven by factors such as differences in infection prevalence and intensity, child age, and intervention duration—we start by following much of the literature and assume that study contexts are exchangeable. We also include fixed-effects models for comparison. Also, since heterogeneity may be driven by variability in worm prevalence, we (1) estimate effects in studies with over 20% and over 50% prevalence only; (2) model the treatment effect on the infected across all studies.

While random-effects estimation of the mean effects of MDA does not rely on distributional assumptions about the true effects, many meta-analyses assume normality or that the distribution is symmetric. For deworming MDA, the distribution of true effects may be non-normal and, in particular, may be right-skewed. This is, first, because deworming drugs are safe and thus are not expected to have negative effects, and second, because most MDA trials have been conducted in low prevalence and low intensity settings, while few trials have been conducted in settings of moderate to high average infection intensity, which account for most of the infection morbidity. Across the MDA settings for which we have data on weight effects, mean worm load is over 6 times larger than median worm load and the Pearson's skewness coefficient is 2.56, suggesting that the distribution of the true effects may also be right-skewed.[13] To address this, we

[13] This is the case assuming that treatment effects are positively correlated to worm load (Anderson, Truscott and Hollingsworth, 2014).

also consider Bayesian estimation of the mean effect of deworming MDA, assuming that the true effects follow distributions that allow for or impose skewness (see Appendix G).

To better characterize the effect of deworming on infected children, one could assume that effects are proportional to worm prevalence and calculate the implied effect of MDA for deworming on infected children based on the estimates from a random effects model of MDA. We do this by dividing the point estimate and the standard error of each study by the reported prevalence of infection in the study population, and using these values in random-effect models.

With pooled estimates of the effect of deworming on infected children across MDA and test-and-treat trials, we next examine whether accounting for differences in infection prevalence may resolve the paradox of positive effects in trials where only infected children are treated, but not in trials including infected and uninfected children. To this end, we conduct random-effect meta-regression of the effects of deworming on infected children.

$$\text{(eqn. 1)} \qquad \hat{\Theta}_{j,k} \sim N\left(a_k + \beta_k MDA_j, \tau_k^2 + \hat{\omega}_{j,k}^2\right)$$

where $\hat{\omega}$ is the standard error of $\hat{\Theta}$, β is the difference in mean impacts between MDA and test-and-treat trials, MDA is an indicator variable for a mass deworming trial, and a is the mean impact of the intervention among test-and-treat trials. We estimate τ_k^2—the cross-trial variance in effect size for outcome k—by the method of moments, which coincides with the DerSimonian and Laird (1986) method when there are no covariates, and focus on the weighted least squares estimator of β_k.

4.3 Cost-effectiveness analysis and test of the hypothesis of non-negative effects across all settings

We next turn to the question of whether the mean nutritional benefits of deworming MDA outweigh its costs. Throughout, we focus on the child nutrition benefits of deworming, ignoring other potential benefits. We take a revealed preference approach to policy choice. First, we search for an intervention that targets similar outcomes in a similar population, and that is widely implemented, suggesting that many policymakers consider the benefits to exceed the costs. We further require that there be a meta-analysis examining the average effect of the policy across settings and that there be data on intervention costs. Second, we compare the expected gains in child nutrition outcomes per $1,000 spent on MDA to those of this alternative intervention. If the gains for MDA are larger for each outcome, and if the policymaker attaches a positive value or weight to each of these gains, then MDA will be more cost-effective than the reference policy for any weighting. Of course, the comparison between MDA and any other intervention will not be perfect, as each intervention might have other effects that we cannot account for in this framework.

We found only one intervention which is implemented in the same populations of preschool and school-aged children and for which comparable cost and meta-analysis estimates of effect are available: preschool and school feeding programs.[14] School feeding is implemented in over 72 countries by the World Food Programme alone (Kristjansson et al., 2007). Kristjansson et al. (2015) examine the impact of preschool feeding programs from 29 different interventions in low- and middle income countries.[15]

[14] Kristjansson et al. (2007) and Kristjansson et al. (2015) conduct Cochrane Reviews of the impact of school and preschool feeding programs, respectively.

[15] We acknowledge that school and preschool feeding programs may not be the most cost-effective nutritional interventions for these populations; however, their wide implementation suggests many policymakers consider the benefits to exceed the costs.

Frequentist random-effects estimates of the mean effects of MDA also have a Bayesian interpretation under some additional assumptions. If (i) a Bayesian policymaker has an uninformative or improper prior about the mean effect of deworming MDA (μ_k), (ii) the cross-trial variance (τ_k) is known and equal to the DerSimonian and Laird (1986) estimate, and (iii) the true effects are normally distributed, then the posterior mean of μ_k corresponds to the random-effects estimate and the posterior variance of μ_k corresponds to the squared standard error. We leverage this equivalence to examine the degree of prior pessimism that a policymaker would need to hold about the effectiveness of MDA such that, after considering all the evidence from MDA trials, the decision-maker would be indifferent between implementing MDA and school feeding programs. We consider a policymaker with a normal prior for μ_k and we define pessimism as the reciprocal of the prior variance (v_k^2), when the prior has a mean of zero.[16] Pessimism measures precision or certainty about the belief of a zero mean effect. For reference, we compare this to the precision obtained when the policymaker has an improper prior (the reciprocal of ${\hat{\sigma}_k^{RE}}^2$).

We also test the hypothesis that deworming MDA has non-negative effects across all settings, against the alternative of a negative effect in at least one setting.[17] We do so within the flexible fixed effects (plural) model (i.e., one that does not assume a common effect), using the likelihood ratio test of qualitative interaction (Gail and Simon, 1985) based on the Q^+ test statistic

(eqn. 2) $$Q_k^+ = \sum_{j=1}^{S_k} \left(\hat{\theta}_{j,k}^2 / \hat{\sigma}_{j,k}^2\right) I\left(\hat{\theta}_{j,k} < 0\right)$$

[16] In this case we note that the posterior mean of μ_k is given by $\bar{\mu}_k = \frac{v_k^2}{\left(\hat{\sigma}_k^{RE}\right)^2 + v_k^2} \hat{\mu}_k^{RE}$

[17] The empirical distribution of estimated effects is a poor guide to this since sampling variation means that, with enough studies, some will yield negative, and even significantly negative, estimates.

where I is an indicator variable.[18] This test will be relevant to a policymaker who would be reluctant to use MDA if there were evidence that MDA yielded negative effects in some settings, e.g., a policymaker who puts larger weight on losses. For completeness, we also test the hypothesis that deworming MDA has non-positive effects across all settings, based on the analogous Q_k^- statistic.[19]

5 Results

5.1 Tests of common zero effect across settings

We begin with two tests of the hypothesis that the true impact of multiple-dose deworming MDA on a given outcome is zero in all settings. We reject the hypothesis that the true impact of deworming MDA is zero in all settings (full sample), against the alternative of a non-zero effect in at least one setting, for weight (p-val<0.001) and MUAC (p-val<0.001), based on the $\sum_{j=1}^{S_k} \hat{\theta}_{j,k}^2 / \hat{\sigma}_{j,k}^2$ test statistic (Table 2, Panel A, row 1). However, we cannot reject the null for height (p-value = 0.18) or hemoglobin (p-value = 0.52). Assuming a common effect of MDA across settings, we reject the hypothesis of a zero effect for weight (p-val<0.001), MUAC (p-val<0.001), and height (p-val=0.048) using the Z test statistic (Table 2, Panel B, row 1). The null of a common zero effect of deworming MDA on hemoglobin is not rejected (p-value = 0.3).[20]

[18] If treatment effect estimates are measured with non-classical error (e.g., due to deviations from protocol, attrition bias) in addition to sampling error, with both sources of error orthogonal to the true effects, this test will reject more often than what the nominal level of the test would suggest.

[19] We report p-values of the Gail and Simon (1985) test as derived in Dmitrienko et al. (2005).

[20] We show the p-values for the test of this hypothesis against the alternative of a common positive effect in Table 2 in square brackets. The results remain qualitatively the same.

We then test the hypothesis that the common effect of deworming MDA on weight, MUAC, height, and hemoglobin are all zero, based on the $\hat{\theta}^{FE'}\Sigma^{-1}\hat{\theta}^{FE}$ test statistic (Table 2, Panel B, rows 2.1-2.3).[21] We reject the null hypothesis that the common effect of deworming MDA on weight, MUAC, height, and hemoglobin are all zero (p-val<0.001).[22]

These results provide substantial evidence against the claim that deworming MDA has a zero effect across settings on all nutrition outcomes. Furthermore, they could directly inform the decision problem of a policymaker who does not believe deworming has negative effects and who believes its costs are negligible.

Results in this subsection remain largely unchanged if one pools MDA and test-and-treat trials (Table 3).

5.2 Estimation of the mean effect of deworming MDA

Following standard approaches in public health and medicine, we conduct both random-effects and fixed-effects estimation of the impact of deworming MDA. In the full sample (of MDA studies) random-effects models we find large values of heterogeneity statistics I^2 and H for models of weight (I^2 = 74%) and MUAC (I^2 = 81%) but not for height (I^2 = 12%) and haemoglobin (I^2 = 1%). Since low estimated heterogeneity may also be due to large sampling variation, below we focus the discussion on the random-effects estimates, but include both fixed- and random-effect estimates in Table 4.

[21] We assume a common correlation coefficient (ρ) across all pairs of fixed-effect estimators, and test the same null for three possible values of : (i) ρ=-0.33, the minimum value for which ∑ is positive semi-definite, (ii) ρ=0, which implies that the fixed-effects estimators of the effect are independent across outcomes, and (iii) ρ=0.99, i.e., nearly perfect correlation across outcomes, the maximum value for which ∑ is positive semi-definite.

[22] We are also able to reject the null hypothesis that the common effect of deworming MDA on weight, MUAC, height, and hemoglobin are all zero for all values of ρ∈-0.33, -0.32,…, 0.98, 0.99, with the maximum p-value being smaller than 0.001.

In the full sample, the estimated mean weight gain effect is 0.140 kg (95% CI: 0.054, 0.227), significant at the 99% confidence level (Table 4, Panel A, column 1). This sample includes trials conducted in low infection prevalence areas where the WHO does not currently recommend mass deworming. In areas with greater than 20% prevalence, where the WHO currently recommends MDA, the estimated mean treatment effect is somewhat larger at 0.154 kg (95% CI: 0.069, 0.240), also significant at the 99% confidence level (Panel C, column 1). We also report weight treatment effects in areas with prevalence greater than 50% (Panel D, column 1), 0.173 kg (95% CI: 0.073, 0.272), and below 20% (Panel B, column 1), 0.112 (95% CI: -0.106, 0.330). As expected, the effect is larger for settings with higher prevalence. The variation is such that we are not able to reject the null hypothesis of zero effect even at the 90% confidence level for settings with prevalence below the WHO's threshold of 20%.

In the full sample, the random-effects estimate of the effect of deworming MDA on MUAC (Panel A, column 3) is 0.127 cm (95% CI: -0.058, 0.313) and the estimate on height (Panel A, column 5) is 0.064 cm (95% CI: -0.018, 0.146). We cannot reject the null hypothesis that the mean effects are zero at conventional levels. However, in settings with over 20% prevalence, the estimated effect on MUAC (Panel C, column 3) is 0.198 cm (95% CI: 0.029, 0.367) and the estimated effect on height (Panel C, column 5) is 0.087 cm (95% CI: 0.011, 0.162), both significantly different from zero at the 95% level. We cannot reject the hypothesis that deworming MDA has a mean zero effect on hemoglobin at the 95% confidence level in either sample (Panels A and C, col 7).[23] Appendix E shows robustness of the result of a significant average effect on weight, MUAC, and height in settings with over 20% prevalence; the weight effect remains

[23] Bayesian estimates of the mean effect of deworming MDA (full sample), based on distributions of true effects that allow for or impose skewness, are always larger than the random-effects estimates (see Appendix G).

significant at the 95% confidence level after dropping any one study estimate and after dropping any pair of estimates, among 210 possible combinations.

The estimated effects of test-and-treat trials (Table 4, Panel E) are positive, significant, and over twice as large as those of MDA trials for weight, MUAC, and height. This should be expected, first, because MDA trials include uninfected children who do not benefit directly from deworming and, second, because average worm load is greater in test-and-treat trials than in MDA trials.[24] When we pool the estimates from MDA and test-and-treat trials (Table 4, Panel F), we also find positive and significant effects of deworming on weight, MUAC, and height. Figures 1-4 show forest plots of the effect of deworming from MDA and test-and-treat trials on weight, MUAC, height, and Hb, respectively.

To contextualize the estimated effects of MDA on weight and height, we compare them to the largest and smallest difference in annual reference weight and height gains by gender (according to WHO growth charts), from birth to age 5, between children at the 15th and 50th percentiles of the respective distribution. The largest difference in annual weight gain between children at the 15th and 50th percentile is 0.6 kg (for boys and girls from birth to age 1); the smallest difference is 0.2 kg (for boys from age 2-3). The estimated MDA treatment effect of 0.154 kg is 26% of the largest annual weight gain gap; and 77% of the smallest gap. For length/height, the largest and smallest 15th-50th percentile annual growth differences are 0.8 cm and 0.4 cm; the estimated treatment effect of 0.087 cm is 11% of the larger (0.8) cm gap and 22% of the smaller (0.4) cm gap.

[24] Earlier versions of the Cochrane Review (Dickson et al., 2000; Dickson et al., 2007) did not create separate categories for test-and-treat and MDA studies as Taylor-Robinson et al. (2015) do, but rather considered all the data together, and the most recent version of the review does not make such distinction either (Taylor-Robinson et al., 2019).

Random-effects estimates based on the full sample indicate that, among infected children, deworming MDA increases child weight by 0.265 kg (p-val=0.004), MUAC by 0.238 cm (p-val=0.043), and height by 0.103 cm (p-val=0.054), see Table 5, Panel A. We also find an average gain in hemoglobin of 0.123 g/dl (p-val=0.100). Deworming of children screened for infection increases weight by 0.657 kg (p-val=0.050), MUAC by 0.396 cm (p-val=0.018), and height by 0.288 cm (p-val=0.061), see Panel B.[25] Combining the evidence of the implied effects of MDA on infected children with the evidence on deworming of screened children, we find that, among infected children, deworming increases weight by 0.327 kg (p-val=0.001), MUAC by 0.272 cm (p-val=0.006), and height by 0.160 cm (p-val=0.010), see Panel C.

The paradox whereby previous meta-analyses find positive effects in studies of treatment of infected children but not in populations including such children could be explained by differences in infection prevalence or intensity. In random-effects meta-regression of these (pooled) effects on an indicator variable for MDA trials, we cannot reject the hypothesis that the mean implied effect of MDA on infected children is the same as the mean effect of deworming of children screened for infection for any of the outcomes (Panel D). Smaller point estimates for effects of deworming of infected children from MDA trials relative to test-and-treat trials are consistent with the fact that the MDA trials took place in settings with lower infection prevalence where infection intensity is also expected to be lower.[26]

[25] These samples correspond exactly to those of the analyses of children known to be infected (single dose) by Taylor-Robinson et al. (2015), excluding Stephenson et al. (1993), which is a misclassified MDA trial (see Section 6).

[26] We also conduct a Wald test of the hypothesis that the mean differences between MDA and test-and-treat trials of the impact of deworming of infected children on weight, MUAC, height, and Hb are jointly zero (see Appendix C).

5.3 Cost-effectiveness analysis and tests of non-negative effects

The estimated gains in child nutrition outcomes per $1,000 spent in deworming treatment are several times larger than those estimated for school and preschool feeding (Table 6). Deworming estimates are based on the random-effects estimates of the effect of deworming MDA on nutrition outcomes, adjusting for the average number of deworming doses, and assuming a cost of $0.68 per person treated for two doses per year.[27] In settings with over 20% worm prevalence, we find that a $1,000 investment in deworming MDA results in nutritional gains of 144.6 kg of weight, 166.5 cm of MUAC, 80.0 cm of height, and 82.6 g/dl of hemoglobin (column 3). Gains are only slightly smaller per $1,000 in the full sample including low prevalence settings (see appendix Table F.2). In settings with over 50% prevalence, the gains per $1,000 investment are 156.1 kg of weight, 166.5 cm of MUAC, 83cm of height, and 65.3 g/dl; they are higher than gains estimated using the 20% threshold for weight and height, equal for MUAC (same studies), and lower for hemoglobin.

Kristjansson et al. (2007) and Kristjansson et al. (2015) conduct Cochrane Reviews of the impact of school and preschool feeding programs, respectively. We combine their estimates of nutritional impact with information on the average duration and costs of these programs (Kristjansson et al., 2016) to estimate the gains in nutrition outcomes per $1,000 spent in school (column 6) and preschool feeding programs (column 9).[28] A $1,000 investment in school feeding programs results in total nutritional gains of 6.2 kg of weight, 3.3 cm of MUAC, and 6.1 cm of height. The estimated gains from deworming MDA in settings with over 20% worm prevalence

[27] This cost estimate is based on data from India (Givewell, 2017) and incorporates the cost of donated drugs, the time that teachers spend administering deworming treatment, among other costs, and thus may be somewhat higher than the costs facing a real-world policymaker.

[28] See Appendix B for details on the cost of school feeding programs, from Kristjansson et al. (2016).

(over 50%) are over 23 (25) times as large for weight, 50 (50) times as large for MUAC, and 13 (13.6) times as large for height.[29] The relative weight and height gains of deworming MDA are similarly large when compared with preschool feeding programs.

The cost-effectiveness of mass deworming is robust to two alternative cost estimates (per person, for two doses). We consider an upper bound cost of \$1.54 (Givewell, 2017) for African countries as well as a lower bound of \$0.38 for India. These upper and lower bound costs are then used to bound estimated gains in child nutrition outcomes per \$1,000 spent in deworming; these are presented in square brackets in Table 6, column 3 (and in Appendix Table F.2).

Leveraging the Bayesian interpretation of the random-effects estimator, for MDA not to be cost-effective relative to school feeding in settings with over 20% infection prevalence, we find that a policymaker would have to believe that the mean weight effect of MDA is zero with an implausible degree of precision, over 22 times as large as the posterior precision obtained with improper priors (Appendix Table F.3). The corresponding factors for MUAC and height effects are 49 and 12 times as large, respectively.

Beyond cost-effectiveness, policymakers who are uncertain about whether deworming drugs have serious side effects could be reluctant to use MDA if there were evidence that MDA could have negative effects in some settings. We test this directly and cannot reject the hypothesis that deworming MDA has non-negative effects across all settings for any of the outcomes, based on the Q^+ test statistic (appendix Table F.1, Panel A, row 1). However, because non-rejection of the null could be driven by limited power, we also report results for the test of the hypothesis that all effects are non-positive, based on the Q^- test statistic (Panel A, row 2). We reject the null that all effects are non-positive for weight (p-val<0.001) and MUAC (p-val<0.001) and reject it for

[29] While the point estimates of the average effect of deworming MDA on hemoglobin are positive, the point estimate of the effect of school feeding on hemoglobin is negative.

height at the 90% confidence level, but cannot reject it for Hb, suggesting that the qualitative interaction tests have less statistical power when examining height and Hb. The results remain qualitatively the same when pooling MDA and test-and-treat trials (appendix Table F.1, Panel B).

6 Revisiting the evidence & interpretation in existing meta-analyses

Here we first present the differences between this paper's sample of MDA trials and that of Taylor-Robinson et al. (2015), and Taylor-Robinson et al. (2019).[30] Second, we demonstrate that the power of the tests implemented by Taylor-Robinson et al. (2015), Taylor-Robinson et al. (2019), and Welch et al. (2016) is such that it cannot rule out a range of meaningful child nutrition effects, including those that would make MDA cost effective relative to school feeding programs.

Appendix Table A.1 shows the differences between this paper's sample and that of Taylor-Robinson et al. (2019) for the analyses of the impact of deworming MDA on weight, MUAC, height, and hemoglobin. The robustness of the findings to these differences in sample construction is examined in appendix Table F.4. Following the release of the review by Taylor-Robinson et al. (2015), we noted that one could obtain a better-powered statistical analysis by, first, including studies that were identified by Taylor-Robinson et al. (2015) but that were excluded from their meta-analysis and, second, by extracting point estimates and standard errors of the impact of deworming MDA using the most precise estimators available. We provided a detailed discussion of these issues in a public working paper version of this study (Croke et al., 2016) and in a formal

[30] We do not contrast this paper's sample to that of the Welch et al. (2016) Campbell Collaboration meta-analysis since the two samples are very similar to one another.

comment submitted to the Cochrane Collaboration.[31] The updated review (Taylor-Robinson et al., 2019) includes several of the additional studies mentioned in the previous version of this paper, but does not include all available trials and estimates, and does not extract data from the most precise estimators available, leading to lower statistical power, and thereby less ability to detect effects of smaller magnitudes. Appendix A provides detailed information on each study in the sample and the differences with Taylor-Robinson et al. (2015, 2019).

In Taylor-Robinson et al. (2019), both the samples and treatment effect estimates are closer to those in this paper than to those of Taylor-Robinson et al. (2015). However, some discrepancies remain. For example, when examining the impact of MDA on weight, Taylor-Robinson et al. (2019) do not emphasize the estimate of an increase of 0.11 kg (95% CI: -0.01, 0.24), which is marginally statistically significant (p-val<0.10). Instead, the authors emphasize a "post-hoc subgroup analysis by studies published prior to and after the year 2000," further noting that "the rationale of the cutpoint was to exclude trials carried out in the previous century, when worm loads were likely to be higher." (pp. 12-13). We argue that emphasizing such a post-hoc analysis might be problematic for the following reasons. First, splitting the sample by whether a trial's (first) article was published before or after the year 2000 is arbitrary; if one is interested in examining effects at different levels of worm load, then it makes more sense to examine this directly, through an analysis of the effects of MDA in settings with different estimated average worm load. Second, while we agree with Taylor-Robinson et al. (2019) that there are fewer higher prevalence settings today than there were in the past, such settings still do exist, and policymakers deciding whether

[31] We submitted this formal comment on Taylor-Robinson et al.'s (2015) review to the Cochrane Collaboration on August 22, 2018, prior to Taylor-Robinson et al.'s (2019) date of last search for trials (September 19, 2018). The comment can be accessed through the link below:
https://drive.google.com/open?id=1P54QA6cI5MbeH8JhbxyolnLRvuQsoNBP

to implement MDA in a high-worm load setting today will find it useful to consider older evidence from settings with comparably high worm load. Third, by effectively dropping half the sample in each of the subgroup analyses, the power to detect a positive effect can be reduced, even if the average effect in a given subgroup is larger than the overall effect, as we show below.

Next, we examine whether the tests of the hypothesis that deworming MDA has a zero average effect on child nutrition, implemented by Taylor-Robinson et al. (2015), Taylor-Robinson et al. (2019), and Welch et al. (2016), were adequately powered to detect effects of a size that would render mass deworming cost-effective relative to feeding programs, for the outcomes analyzed in this paper.[32] Table A.2 reproduces the estimates of the average child nutrition effects of deworming MDA from the main analysis by Taylor-Robinson et al. (2015), Taylor-Robinson et al. (2019), and Welch et al. (2016). Taylor-Robinson et al. (2015) report random-effects estimates for weight, MUAC, and hemoglobin, and a fixed-effects estimate for height. Taylor-Robinson et al. (2019) report random-effects estimates for weight and MUAC, and fixed-effects estimates for height and hemoglobin.[33] Welch et al. (2016) report random-effects estimates in terms of standardized mean differences rather than kg (for weight) or cm (for height) so that they can combine, in a single specification, studies using different outcomes (for example, weight in kg and weight-for-age z scores); they do not report point estimates for MUAC or hemoglobin. Based on these estimates, the respective authors tested the hypothesis that deworming MDA has a zero average effect on each outcome. As Table A.2 shows, neither Taylor-Robinson et al. (2015),

[32] While Taylor-Robinson et al. (2019) broadened the main category of analysis of their updated review to include multiple-dose trials that screened children for infection, it remained de facto the same as they were unable to identify any such trial.

[33] For the post-hoc analyses, Taylor-Robinson et al. (2019) present random-effects estimates both for the pre and post 2000 samples. For the latter analysis, the estimate of the cross-trial variance is zero and, therefore, the random-effects estimate is equivalent to the fixed-effects estimate.

Taylor-Robinson et al. (2019), nor Welch et al. (2016) reject the null for any of their outcomes, at the conventional 95% confidence level.

To examine whether these tests are adequately powered, we first calculate the minimum detectable effect (MDE) to reject the null hypothesis of a zero average effect at the 95% confidence level, with 80% power (panel B).[34] The minimum detectable effect for the main analysis of weight gain in Taylor-Robinson et al. (2015) is 0.276 kg; this MDE is reduced in the update (Taylor-Robinson et al., 2019) to 0.181 kg (partly as a result of the inclusion of some previously omitted trials), and the MDE for Welch et al. (2016) is 0.294 kg.[35] We also calculate the minimum average effect that renders deworming cost-effective relative to school and preschool feeding programs (panel C).[36] The MDEs in these studies are orders of magnitude larger than the minimum effect that renders deworming cost-effect relative to feeding programs, implying that these tests lack power to reject effects that would make deworming MDA a desirable policy option relative to other popular policies aimed at improving child nutrition in similar populations.[37] Note that

[34] These estimates were obtained using the method of Hedges and Pigott (2001). In particular, for a given effect size, we estimate power as: Power= 1−Φ(1.96-EffectSize/StandardError)+Φ(-1.96-EffectSize/StandardError), where Φ is the cumulative distribution function for a standard normal random variable, and StandardError is the standard error for the average effect size under the random effects model. Reported power for a given effect size is the probability that the null hypothesis that the average effect size is zero is rejected at the 0.05 level of significance. The reported MDE is an estimate of the effect size that would deliver a test with 80% power

[35] For comparison, in settings with over 20% infection prevalence, our MDE is 0.122 kg for weight, 0.242 cm for MUAC, 0.108 cm for height, and 0.108 g/dl for Hb.

[36] These effects are calculated as the product of the outcome gain per dollar spent in school or preschool feeding programs and the average cost of deworming MDA, which is calculated as the product of the cost per deworming treatment ($0.34) and the average number of doses across trials.

[37] The implicit loss function implied by requiring 95% confidence to undertake MDA without regard to the statistical power of the test is one in which there is a high cost of a false positive and a low cost of a false negative. That might be appropriate if, for example, the US Food and Drug Administration were considering a drug that might have major side effects or very high costs. However deworming drugs have already been through regulatory approval and the monetary cost of deworming is low, while there is some evidence that deworming has large long-run benefits (Ahuja et al. 2015). Thus, the cost of a false positive is low while the cost of a false negative is potentially substantial in endemic areas. In these situations, policymakers following the decision rule implied by a frequentist test may achieve higher welfare levels by using lower significance level thresholds, reducing the probability of incurring type II error, while incurring a greater probability of low-cost type I error (Manski, 2007).

Taylor-Robinson et al.'s (2019) post-hoc analysis for trials published before the year 2000 has an MDE which is close to twice the size of the impact that they estimate (i.e., an increase of 0.258 kg), indicating that this analysis is also substantially underpowered.[38]

In Appendix D, we show that the meta-analysis by Welch et al. (2016) is underpowered primarily because they subdivide deworming studies based on the type of drugs used, the frequency of treatment, and whether the trial compared deworming to pure placebo versus trials in which deworming plus an additional intervention is compared to the additional intervention alone. Thus, instead of reporting a single meta-analysis which aggregates a large number of studies, they conduct multiple small-sample meta-analyses. When one relaxes the narrow category of analysis from the main comparison in Welch et al. (2016) – for example, to include trials where approved drugs aside from albendazole were used, or where deworming was done more or less frequently than twice per year – one obtains statistically significant estimates of the effect of deworming on weight, and in some cases for height. Results are shown in Appendix D, where we also explore other reasons why Welch et al. (2016) has limited statistical power.

In Figure 5 we compare the estimates (and confidence intervals) of the mean effects of deworming MDA on weight and height from Taylor-Robinson et al. (2015), Taylor-Robinson et al. (2019), and Welch et al. (2016), relative to the ones we estimate here. For both outcomes, the larger sample used in this study results in more precise estimators of the mean effect of deworming MDA. This graph also indicates a convergence in results across different review groups. The initial

[38] The weight MDE from the analysis of trials published from the year 2000 onwards (0.083 kg) is smaller than that from the analysis of trials published before the year 2000 (0.461 kg). However, even the smaller MDE is an order of magnitude larger than the effect that would render MDA cost-effective relative to school feeding (0.009 kg). That the MDE in the former analysis is smaller than in the latter is partly due to the fact that the estimated cross-trial variance in the post-2000 analysis is zero, which mechanically leads to increased precision in the estimator. However, this parameter could be biased towards zero if trials in high-prevalence settings were omitted from the analysis. We note that Taylor-Robinson et al. (2019) exclude Carmona-Fonseca et al. (2015) and Wiria et al. (2013), both in settings with over 20% prevalence.

estimates of the weight and height effects of MDA from Taylor-Robinson et al. (2015) were small and had wide confidence intervals around them. Welch et al. (2016) and Taylor-Robinson et al. (2019) incorporate additional trials, but sacrifice power in other ways (e.g., splitting samples, not obtaining estimates from most precise estimators), obtaining larger point estimates with tighter confidence intervals—but effects they estimate are still not statistically significant at the 95% level. While we obtain point estimates that are close to those estimated by Welch et al. (2016) and Taylor-Robinson et al. (2019), the precision of our estimators is improved by addressing the issues above. As a consequence, we estimate statistically significant impacts and, as expected, these are larger in settings where the WHO recommends MDA.

7 Conclusion

We report on a meta-analysis of the nutritional impact of deworming. Recent meta-analyses fail to reject the hypothesis that deworming MDA has a zero mean effect on child nutrition outcomes and argue that MDA is ineffective and should be discontinued (Taylor-Robinson et al., 2015; Taylor-Robinson et al., 2019; Welch et al., 2016). They advocate this policy change despite finding positive effects across several nutritional outcomes from deworming of children known to be infected. This creates a paradox: if deworming has a positive effect on infected individuals, one would expect a smaller but positive effect from MDA in endemic populations. We show that these studies are underpowered to detect effects that would render MDA cost-effective relative to a relevant alternative policy of school-feeding. In addition, by splitting their samples into different categories of analysis or excluding relevant trials, these studies sacrifice statistical power.

We tested a series of hypotheses aimed at informing the deworming question. To this end, we make use of the most comprehensive set of trials examining the effects of deworming MDA on children’s nutrition, which allows us to conduct higher powered statistical tests than previous studies. When aggregating all available estimates of the effect of MDA deworming, we first reject the null hypothesis that deworming has a common zero effect for weight, MUAC, and height, and reject the hypothesis that the (common) effects of the four outcomes examined are jointly zero. We then estimate the mean effect of deworming MDA. In areas where the WHO recommends MDA (>20% and >50% prevalence), we find that multiple-dose deworming significantly increases child weight, MUAC, and height, and that MDA is many times more cost-effective than widely implemented school-feeding programs. A Bayesian analysis suggests that policymakers would need extremely confident priors that MDA has no effect in order not to believe that it is more cost effective than school feeding.

Next, motivated by the paradox of large effects from treatment of infected children but not in MDA trials, we compare the implied treatment effect on infected children that participated in MDA trials versus the mean effect on screened children in test-and-treat trials. We cannot reject the hypothesis that the mean effect is the same between MDA and test-and-treat trials for any outcomes. This helps resolve the apparent paradox and is compatible with the hypothesis that the lower average infection prevalence and intensity in MDA trials compared to test-and-treat trials could lead meta-analyses to be underpowered for the detection of their effects. We hope that future work may build on this finding by explicitly investigating the relationship between infection intensity and deworming effects.

The estimates in the sample are likely lower bounds of the effects that would be obtained from treating entire endemic populations, because the studies in this literature generally do not address epidemiological externalities (Miguel and Kremer, 2004; Bundy et al., 1990). Most trials in the sample were randomized at the individual level and, even when trials are randomized at the cluster-level, no study with the exception of Miguel and Kremer (2004) estimates the potential epidemiological spillovers. Therefore, this paper's estimates of the average effect of MDA are likely also lower bounds. The finding that deworming improves nutrition in at least some settings implies that the literature on the long-run educational and economic impacts of deworming cannot be dismissed a priori; that literature suggests that the expected long-run benefits of mass deworming greatly exceed the cost.

We argue that policy choice benefits from a decision theory perspective. While the standard approach to meta-analysis focuses on the question of *whether* MDA has a zero average effect, we argue that the most pressing policy question in the case of deworming MDA is rather *where* MDA can be expected to be cost-effective. On the one hand, there is a consensus in the public health community that infected children should be treated and it is uncontroversial to treat very high prevalence populations. On the other hand, there is no question that worm-free populations, or those with very low (e.g., 1%) infection prevalence, should not receive MDA. While there is uncertainty about the optimal threshold of infection prevalence or intensity that would warrant deworming MDA, at minimum it is evident that MDA generates nutritional gains for children in some circumstances, with larger estimated gains in settings with more infections as would be expected. This is supported by findings in this paper, which shows that MDA has positive effects in settings with over 20% prevalence—the threshold endorsed by the WHO—and is substantially more cost-effective than a leading alternative nutritional intervention.

References

Ahuja, Amrita, Sarah Baird, Joan Hamory Hicks, Michael Kremer, Edward Miguel and Shawn Powers. 2015. "When Should Governments Subsidize Health? The Case of Mass Deworming." The World Bank Economic Review 29(suppl 1):S9–S24.

Albonico, M., Allen, H., Chitsulo, L., Engels, D., Gabrielli, A.F. and Savioli, L., 2008. Controlling soil-transmitted helminthiasis in pre-school-age children through preventive chemotherapy. PLoS neglected tropical diseases, 2(3), p.e126.

Anderson, R, J Truscott and TD Hollingsworth. 2014. "The coverage and frequency of mass drug administration required to eliminate persistent transmission of soil-transmitted helminths." Phil. Trans. R. Soc. B 369:20130435

Anderson, R.M., Truscott, J.E., Pullan, R.L., Brooker, S.J. and Hollingsworth, T.D., 2013. How effective is school-based deworming for the community-wide control of soil-transmitted helminths?. PLoS neglected tropical diseases, 7(2), p.e2027.

Andrews, Isaiah. and Maximilian Kasy. 2019. Identification of and correction for publication bias. American Economic Review 109(8):2766-94.

Anderson, RM and RM May. 1985. "Helminth infections of humans: mathematical models, population dynamics, and control." Adv Parasitol 24:1–101.

Assefa, L. M., T. Crellen, S. Kepha, J. H. Kihara, S. M. Njenga, R. L. Pullan and S. J. Brooker. 2014. "Diagnostic Accuracy and Cost-Effectiveness of Alternative Methods for Detection of Soil-Transmitted Helminths in a Post-Treatment Setting in Western Kenya." PLoS Neglected Tropical Diseases 8(5):e2843.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward. 2016. "Worms at Work: Long-Run Impacts of a Child Health Investment" The Quarterly Journal of Economics 131(4): 1637-1680. doi: 10.1093/qje/qjw022

Barda, B., H, Zepherine, L. Rinaldi, G. Cringoli, R. Burioni, M. Clementi and M. Albonico. 2013. "Mini-FLOTAC and Kato-Katz: Helminth Eggs Watching on the Shore of Lake Victoria." Parasites & Vectors 6(220).

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

Bleakley, Hoyt. 2018. "Report on "Disease and Development [...] Comment"." Manuscript.

Bundy, D.A., Wong, M.S., Lewis, L.L. and Horton, J., 1990. Control of geohelminths by delivery of targeted chemotherapy through schools. Transactions of the Royal Society of Tropical Medicine and Hygiene, 84(1), pp.115-120.

Carmona-Fonseca, Jaime, and Adriana Correa-Botero. "Efecto del albendazol y la vitamina A periódicos sobre helmintos intestinales y anemia en niños del Urabá Antioqueño (Colombia)." Biosalud 14.1 (2015): 9-25.

Croke, K. and Atun, R., 2019. The long run impact of early childhood deworming on numeracy and literacy: Evidence from Uganda. PLoS neglected tropical diseases, 13(1), p.e0007085.

Dahal, M. and Fiala, N., 2018. What do we know about the impact of microfinance? The problems of power and precision (No. 756). Ruhr Economic Papers.

DerSimonian, R. and Laird, N., 1986. Meta-analysis in clinical trials. Controlled clinical trials, 7(3), pp.177-188.

Dickson, R, S Awasthi, C Demellweek and P Williamson. 2000. "Anthelmintic drugs for treating worms in children: effects on growth and cognitive performance (Review)." The Cochrane Library (2).

Dickson, R, S Awasthi, C Demellweek and P Williamson. 2007. "Anthelmintic drugs for treating worms in children: effects on growth and cognitive performance (Review)." The Cochrane Library (2).

Dickson, R, S Awasthi and C Demmellweek. 1997. "Routine Intermittent Anthelminth Therapy in Disadvantaged Populations [Protocol].".

Dmitrienko, A., Molenberghs, G., Chuang-Stein, C. and Offen, W.W., 2005. Analysis of clinical trials using SAS: A practical guide. SAS Institute.

Gail, M. and Simon, R., 1985. Testing for qualitative interactions between treatment effects and patient subsets. Biometrics, pp.361-372.

Garner, Paul, David Taylor-Robinson and Harshpal Singh Sachdev. 2015. "Commentary: Replication of influential trial helps international policy." International Journal of Epidemiology 44(5):1599–1601.

Givewell. 2011. URL: http://blog.givewell.org/2011/09/29/errors-in-dcp2-cost-effectiveness-estimate-fordeworming/#actualcosteffectivenes

Givewell. 2014. URL: http://www.givewell.org/international/top-charities/deworm-world-initiative

Givewell. 2017. "Evidence Action's Deworm the World Initiative".
URL: https://www.givewell.org/charities/deworm-world-initiative#What_is_the_cost_per_treatment

Hadju, V., Stephenson, L.S., Abadi, K., Mohammed, H.O., Bowman, D.D. and Parker, R.S., 1996. Improvements in appetite and growth in helminth-infected schoolboys three and seven weeks after a single dose of pyrantel pamoate. Parasitology, 113(5), pp.497-504.

Hall, Andrew and Sue Horton. 2008. Best Practice Paper: Deworming. Copenhagen Consensus Center, Denmark.

Hamory, Joan, Edward Miguel, Michael Walker, Michael Kremer, and Sarah Baird. 2021. Twenty Year Economic Impacts of Deworming. Proceedings of the National Academy of Sciences, 118(14): e2023185118.

Higgins, J., Thompson, S.G. and Spiegelhalter, D.J., 2009. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 172(1), pp.137-159.

Higgins, Julian PT and Sally Green. 2011. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane.

Horton, J. 2000. “Albendazole: A Review of Anthelmintic Efficacy and Safety in Humans.” Parasitology 121((Suppl)):S11332.

Hotez, Peter J, Donald A. P. Bundy, Kathleen Beegle, Simon Brooker, Lesley Drake, Nilanthi de Silva and Lorenzo Savioli. 2006. Helminth Infections: Soil-transmitted Helminth Infections and Schistosomiasis, from Disease Control Priorities in Developing Countries (2nd edition). World Bank.

Ioannidis, J.P.A., T.D. Stanley, and H. Doucouliagos. 2017. “The Power of Bias in Economics Research”, Economic Journal, 127(605): F236-F265, 10.1111/ecoj.12461.

Jinabhai, C.C., Taylor, M., Coutsoudis, A., Coovadia, H.M., Tomkins, A.M. and Sullivan, K.R., 2001. Epidemiology of helminth infections: implications for parasite control programmes, a South African perspective. Public health nutrition, 4(6), pp.1211-1219.

J-PAL Policy Bulletin. 2012. “Deworming: A Best Buy for Development.”.
URL: https://www.povertyactionlab.org/sites/default/files/publications/2012.3.22-Deworming.pdf

Joseph, Serene A., Martin Casapia, Antonio Montresor, Elham Rahme, Brian J. Ward, Grace S. Marquis, Lidsky Pezo, Brittany Blouin, Mathieu Maheu-Giroux and TheresaW. Gyorkos. 2015. “The Effect of Deworming on Growth in One-Year-Old Children Living in a Soil-Transmitted Helminth-Endemic Area of Peru: A Randomized Controlled Trial.” PLoS Negl Trop Dis 9.

Joseph, S.A., Montresor, A., Casapía, M., Pezo, L. and Gyorkos, T.W., 2016. Adverse events from a randomized, multi-arm, placebo-controlled trial of mebendazole in children 12–24 months of age. The American journal of tropical medicine and hygiene, 95(1), pp.83-87.

Keiser, J. and J. Utzinger. 2008. “Efficacy of Current Drugs against Soil-Transmitted Helminth Infections: Systematic Review and Meta-Analysis.” Journal of the American Medical Association 299(16):193748.

Koroma, M.M., Williams, R.A.M., De La Haye, R. and Hodges, M., 1996. Effects of albendazole on growth of primary school children and the prevalence and intensity of soil-transmitted helminths in Sierra Leone. Journal of tropical pediatrics, 42(6), pp.371-372.

Kristjansson, E. A., V. Robinson, M. Petticrew, B. MacDonald, J. Krasevec, L. Janzen, T. Greenhalgh, G.Wells, J. MacGowan, A. Farmer, B. J. Shea, A. Mayhew and P. Tugwell. 2007. “School feeding for improving the physical and psychosocial health of disadvantaged elementary school children.” Cochrane Database Syst Rev (1):CD004676.

Kristjansson, E., Francis, D.K., Liberato, S., Benkhalti Jandu, M., Welch, V., Batal, M., Greenhalgh, T., Rader, T., Noonan, E., Shea, B. and Janzen, L., 2015. Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years. The Cochrane Library.

Kristjansson, E.A., Gelli, A., Welch, V., Greenhalgh, T., Liberato, S., Francis, D. and Espejo, F., 2016. Costs, and cost-outcome of school feeding programmes and feeding programmes for young children. Evidence and recommendations. International Journal of Educational Development, 48, pp.79-83.

Lo, N. C., Y. S. Lai, D. A. Karagiannis-Voules, I. I. Bogoch, J. T. Coulibaly, E. Bendavid, J. Utzinger, P. Vounatsou and J. R. Andrews. 2016. “Assessment of global guidelines for preventive chemotherapy against schistosomiasis and soil-transmitted helminthiasis: a cost-effectiveness modelling study.” Lancet Infect Disease.

Manski, C.F., 2007. Adaptive minimax-regret treatment choice, with application to drug approval (No. w13312). National Bureau of Economic Research.

McKenzie, David. 2012. “Beyond baseline and follow-up: The case for more T in experiments.” Journal of Development Economics 99(2):210 – 221.

Meager, Rachael. 2018. Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments. American Economic Journal: Applied Economics.

Meager, Rachael. 2020. Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Approach to the Microcredit Literature. American Economic Review (forthcoming).

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

Miguel, Edward and Michael Kremer. 2014. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities, Guide to Replication of Miguel and Kremer (2004).”
URL: http://emiguel.econ.berkeley.edu/assets/miguel research/46/PSDP-REP 2014-11.pdf

Ostwald, Rosemarie, Mark Fitch, Rainer Arnhold, Jennifer Shield, Louie Dexter, Jan Kilner and Richard Kimber. 1984. “The effect of intestinal parasites on nutritional status in well-nourished school age children in Papua New Guinea.” Nutrition Reports International 30(6):1409–1421.

Ozier, Owen. 2018. “Exploiting Externalities to Estimate the Long-term Benefits of Early Childhood Deworming.” American Economic Review: Applied Economics 10(3): 235-262.

Palupi, L., Schultink, W., Achadi, E. and Gross, R., 1997. Effective community intervention to improve hemoglobin status in preschoolers receiving once-weekly iron supplementation. The American journal of clinical nutrition, 65(4), pp.1057-1061.

Perez del Villar, L., F. J. Burguillo, J. Lopez-Aban and A. Muro. 2012. “Systematic Review and Meta Analysis of Artemisinin Based Therapies for the Treatment and Prevention of Schistosomiasis.” PLoS ONE 7(9):e45867.

Pullan, R. L., J. L. Smith, R. Jasrasaria and S. J. Brooker. 2014. “Global Numbers of Infection and Disease Burden of Soil Transmitted Helminth Infections in 2010.” Parasites and Vectors 7(37).

Rice, K., Higgins, J. and Lumley, T., 2018. A re-evaluation of fixed effect(s) meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 181(1), pp. 205-227.

Rohatgi, Ankit. 2015. “WebPlotDigitizer.” URL: http://arohatgi.info/WebPlotDigitizer

Roodman, David. 2017a. “Comment: The impacts of the hookworm eradication in the American South.” URL: http://files.givewell.org/files/droodman/Bleakley-replication-9.pdf

Roodman, David. 2017b. How thin the reed? Generalizing from “Worms at Work”
URL: https://blog.givewell.org/2017/01/04/how-thin-the-reed-generalizing-from-worms-at-work/

Stephenson, L. S., M. C. Latham, E. J. Adams, S. N. Kinoti and A. Pertet. 1993. “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.” Journal of Nutrition 123(6):103646.

Stephenson, LS, MC Latham, KM Kurz, SN Kinoti and H. Brigham. 1989. “Treatment with a single dose of albendazole improves growth of Kenyan schoolchildren with hookworm, Trichuris trichiura, and Ascaris lumbricoides infections.” American Journal of Tropical Medicine and Hygiene 41(1).

Taylor-Robinson, D.C., Maayan, N., Donegan, S., Chaplin, M. and Garner, P., 2019. Public health deworming programmes for soil-transmitted helminths in children living in endemic areas. Cochrane Database of Systematic Reviews, (9).

Taylor-Robinson, D.C., Maayan, N., Soares-Weiser, K., Donegan, S. and Garner, P., 2015. Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin, and school performance. Cochrane Database of Systematic Reviews, (7).

Taylor-Robinson, DC, N Maayan, K Soares-Weiser, S Donegan and P Garner. 2012. "Deworming drugs for soil-transmitted intestinal Deworming drugs for soil-transmitted intestinal worms in children: effects on nutritional indicators, haemoglobin and school performance (Review)." The Cochrane Library (11).

Turner, R. M., S. M. Bird and J. P. Higgins. 2013. "The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews." PLoS ONE 8(3):e59202.

Vivalt, Eva. 2015. "Heterogeneous Treatment Effects in Impact Evaluation." American Economic Review, Papers and Proceedings, 105(5):467–70.

Vivalt, Eva. 2020. "How Much Can We Generalize From Impact Evaluations?" Journal of the European Economic Association 18(6): 3045–3089.

Watkins, William E., Jos R. Cruz and Ernesto Pollitt. 1996. "The effects of deworming on indicators of school performance in Guatemala." Transactions of The Royal Society of Tropical Medicine and Hygiene 90(2):156–161.

Welch, V.A., Awasthi, S., Cumberbatch, C., Fletcher, R., McGowan, J., Merritt, K., Krishnaratne, S., Sohani, S., Tugwell, P. and Wells, G.A., 2016. Deworming and adjuvant interventions for improving the developmental health and well-being of children in low-and middle-income countries. Campbell Systematic Reviews, 12.

World Bank. 1993. World Development Report 1993: Investing in Health. World Bank.

World Health Organization. 2002. "Prevention and Control of Schistosomiasis and Soil-Transmitted Helminthiasis." URL: http://apps.who.int/iris/bitstream/10665/42588/1/WHO TRS 912.pdf

World Health Organization. 2017. Guideline: preventive chemotherapy to control soil-transmitted helminth infections in at-risk population groups.
URL: http://apps.who.int/iris/bitstream/10665/258983/1/9789241550116-eng.pdf

Wooldridge, J.M., 2010. Econometric analysis of cross section and panel data. MIT Press.

Table 1: Summary of treatment effects and prevalence of worms in included studies

Study	Treatment Effects (Standard errors)				Worm prevalence*
	Weight (kg)	Height (cm)	Mid-Upper Arm Circumference (cm)	Hemoglobin (g/dL)	(%)
Panel A: MDA trails					
Alderman 2006	0.154				76
	(0.089)				
Awasthi 1995	0.980	1.190			8
	(0.148)	(1.204)			
Awasthi 2000	-0.050	-0.410		0	12
	(0.076)	(0.314)		(0.041)	
Awasthi 2001	0.170	0.400			9
	(0.065)	(0.31)			
Carmona-Fonseca 2015a	0.201	-0.067		0.129	45
	(0.136)	(0.193)		(0.091)	
Carmona-Fonseca 2015b	0.062	-0.067		0.007	45
	(0.118)	(0.193)		(0.082)	
Donnen 1998	-0.450	-1.190	-0.35		10
	(0.166)	(0.552)	(0.154)		
Dossa 2001a	0	0.500	0	0.3	58
	(0.265)	(0.637)	(0.215)	(0.299)	
Dossa 2001b	0	0	0.1	0.2	58
	(0.138)	(0.317)	(0.188)	(0.329)	
Gateff 1972	0.347				76
	(0.13)				
Gupta 1982a	0.027	-0.095			62
	(0.175)	(0.444)			
Gupta 1982b	0.130	-0.029			59
	(0.148)	(0.474)			

Hall 2006	0.054	0.089	0.794		84
	(0.058)	(0.082)	(0.314)		
Joseph 2015	0.040	0.040			11
	(0.049)	(0.127)			
Kirwan 2010				0.170	46
				(0.121)	
Kruger 1996	-0.380	0.080		-0.020	38
	(0.226)	(0.21)		(0.154)	
Kruger 1996b	0.393	0.209		0.269	38
	(0.186)	(0.208)		(0.129)	
Le Huong 2007a				-0.080	73
				(0.136)	
Le Huong 2007b				0.030	73
				(0.129)	
Liu 2017	0.030	0.080		-0.043	31
	(0.127)	(0.354)		(0.108)	
Miguel 2004	-0.618				77
	(0.304)				
Ndibazza 2012	0.010	-0.230		-0.070	3
	(0.091)	(0.285)		(0.063)	
Ostwald 1984	0.700	0.300		0.300	92
	(0.449)	(0.27)		(0.277)	
Rousham 1994			0.1		71
			(0.058)		
Stephenson 1993	0.900	-0.100	0.400		88
	(0.184)	(0.163)	(0.064)		
Stoltzfus 1997a	0.234	0.218			95
	(0.098)	(0.086)			
Stoltzfus 1997b	0.110	0			95

	(0.139)	(0.098)			
Sur 2005	0.292				53
	(0.088)				
Watkins 1996	0.130	0.060	0.080		91
	(0.106)	(0.098)	(0.07)		
Willett 1979	0.160				53
	(0.085)				
Wiria 2013	0.188	1.348			76
	(0.394)	(0.535)			
Panel B: Test-and-treat trials					
Freij 1979a	0.200		-0.300		100
	(1.47)		(0.713)		
Freij 1979b			0.100		100
			(0.347)		
Sarkar 2002	0.380	0.100			100
	(0.15)	(0.261)			
Stephenson 1989	1.300	0.600	0.500		97
	(0.134)	(0.134)	(0.078)		
Tee 2013		-0.100			15
		(0.404)			
Yap 2014	0.300	0.200		-0.4	96
	(0.179)	(0.128)		(0.434)	

*Note: For each study, the prevalence variables is defined as the maximum of prevalences over all worms reported in the study

Table 2: Testing for the existence of child nutrition impacts of mass deworming

	Weight (W)	MUAC (M)	Height (H)	Hemoglobin (Hb)	W, M, H, Hb (jointly)
	1	2	3	4	5
Panel A: Allowing for treatment heterogeneity					
1. H_0: All treatment effects are zero (p-value)					
	<0.001	<0.001	0.181	0.524	
Panel B: Assuming a common treatment effect					
1. H_0: Common effect is zero (p-value)†					
	<0.001	<0.001	0.048	0.296	
	[<0.001]	[<0.001]	[0.024]	[0.148]	
2. H_0: Common effects are zero (p-value)					
2.1				ρ=-0.33:	<0.001
2.2				ρ=0:	<0.001
2.3				ρ=0.99:	<0.001

Notes: The test statistics used for testing these hypotheses are described in subsection 4.1. The sample is the full set of MDA trials. †The p-value of the one-tailed test of the hypothesis of a zero common effect against the alternative of a positive common effect is presented in square brackets. ρ is the assumed correlation coefficient between any pair of fixed-effect estimators: ρ=-0.33 and ρ=0.99 are the minimum and maximum values of ρ for which the variance-covariance matrix of the vector of fixed-effect estimators is positive semi-definite, respectively.

Table 3: Testing for the existence of child nutrition impacts of deworming combining MDA and test-and-treat trials

	Weight (W)	MUAC (M)	Height (H)	Hemoglobin (Hb)	W, M, H, Hb (jointly)
	1	2	3	4	5
Panel A: Allowing for treatment heterogeneity					
1. H0: All treatment effects are zero (p-value)					
	<0.001	<0.001	0.003	0.534	
Panel B: Assuming a common treatment effect					
1. H0: Common effect is zero (p-value)†					
	<0.001	<0.001	<0.001	0.324	
	[<0.001]	[<0.001]	[<0.001]	[0.162]	
2. H0: Common effects are zero (p-value)					
2.1				ρ=-0.33:	<0.001
2.2				ρ=0:	<0.001
2.3				ρ=0.99:	<0.001

Notes: The test statistics used for testing these hypotheses are described in subsection 4.1. The sample is the full set of MDA trials and test-and-treat trials. †The p-value of the one-tailed test of the hypothesis of a zero common effect against the alternative of a positive common effect is presented in square brackets. ρ is the assumed correlation coefficient between any pair of fixed-effect estimators: ρ=-0.33 and ρ=0.99 are the minimum and maximum values of ρ for which the variance-covariance matrix of the vector of fixed-effect estimators is positive semi-definite, respectively.

Table 4: Random-effects and fixed-effect estimates

	Weight (kg)		MUAC (cm)		Height (cm)		Hb (g/dl)	
	1	2	3	4	5	6	7	8
Est. Method	RE	FE	RE	FE	RE	FE	RE	FE
Panel A: Full Sample of MDA trials								
Point estimate	0.140	0.117	0.127	0.164	0.064	0.071	0.028	0.028
Se.	0.044	0.020	0.095	0.034	0.042	0.036	0.027	0.027
p-val†	0.002	<0.001	0.180	<0.001	0.124	0.048	0.296	0.296
	[<0.001]	[<0.001]	[0.090]	[<0.001]	[0.062]	[0.024]	[0.148]	[0.148]
	N=27		N=7		N=22		N=13	
Panel B: MDA trials with <20% prevalence								
Point estimate	0.112	0.076	-0.35	-0.35	-0.108	-0.035	-0.011	0.011
Se.	0.111	0.031	0.154	0.154	0.181	0.101	0.038	0.038
p-val†	0.314	0.015	0.023	0.023	0.548	0.729	0.773	0.773
	[0.157]	[0.007]	[0.988]	[0.988]	[0.726]	[0.636]	[0.614]	[0.614]
	N=6		N=1		N=6		N=2	
Panel C: MDA trials with ≥20% prevalence								
Point estimate	0.154	0.147	0.198	0.191	0.087	0.087	0.069	0.069
Se.	0.044	0.027	0.086	0.035	0.038	0.038	0.038	0.038
p-val†	<0.001	<0.001	0.022	<0.001	0.024	0.024	0.073	0.073
	[<0.001]	[<0.001]	[0.011]	[<0.001]	[0.012]	[0.012]	[0.037]	[0.037]
	N=21		N=6		N=16		N=11	
Panel D: MDA trials with ≥50% prevalence								
Point estimate	0.173	0.157	0.198	0.191	0.095	0.096	0.044	0.044
Se.	0.051	0.029	0.086	0.035	0.048	0.042	0.082	0.082
p-val†	<0.001	<0.001	0.022	<0.001	0.049	0.022	0.589	0.589
	[<0.001]	[<0.001]	[0.011]	[<0.001]	[0.025]	[0.011]	[0.295]	[0.295]
	N=16		N=6		N=11		N=5	
Panel E: Test-and-treat trials								
Point estimate	0.646	0.748	0.401	0.472	0.287	0.337	-0.400	-0.400
Se.	0.325	0.087	0.152	0.076	0.149	0.085	0.434	0.434
p-val†	0.047	<0.001	0.008	<0.001	0.054	<0.001	0.356	0.356
	[0.023]	[<0.001]	[0.004]	[<0.001]	[0.027]	[<0.001]	[0.822]	[0.822]

	N=4		N=3		N=4		N=1	
Panel F: Pooling all MDA and test-and-treat trials								
Point estimate	0.194	0.150	0.174	0.217	0.102	0.111	0.026	0.026
Se.	0.053	0.020	0.089	0.031	0.048	0.033	0.027	0.027
p-val†	<0.001	<0.001	0.051	<0.001	0.035	<0.001	0.324	0.324
	[<0.001]	[<0.001]	[0.025]	[<0.001]	[0.018]	[<0.001]	[0.162]	[0.162]
	N=31		N=10		N=26		N=14	

Notes: Estimation method is random-effects (RE) in odd numbered columns and fixed-effect (FE) in even numbered columns. †The p-value of the one-tailed test of the hypothesis of no effect against the alternative of a positive effect is presented in square brackets. The random-effects and fixed-effect estimates for the height and hemoglobin effects, in settings with over 20% worm prevalence (Panel B), are nearly identical (identical up to three decimal points) given that the estimated between-trial variances are small: 0.0043 for height and 0.0001 for hemoglobin. In the case of hemoglobin, the two estimates are also nearly identical in the other settings.

Table 5: Effects of deworming of infected children

	Weight (kg)	MUAC (cm)	Height (cm)	Hb (g/dl)
	1	2	3	4
Panel A: MDA trials				
RE estimate	0.265	0.238	0.103	0.123
s.e.	0.091	0.118	0.053	0.075
p-val	0.004	0.043	0.054	0.100
N	27	7	22	13
Panel B: Test-and-treat trials				
RE estimate	0.657	0.396	0.288	-0.400
s.e.	0.336	0.167	0.154	0.434
p-val	0.050	0.018	0.061	0.356
N	4	3	4	1
Panel C: MDA and test-and-treat trials				
RE estimate	0.327	0.272	0.160	0.108
s.e.	0.096	0.099	0.062	0.074
p-val	0.001	0.006	0.010	0.143
N	31	10	26	14
Panel D: Test of the hypothesis that the average effect of deworming of infected children is the same between MDA and test-and-treat trials				
Difference	-0.407	-0.127	-0.220	0.523
s.e.	0.251	0.228	0.127	0.440
p-val	0.105	0.577	0.083	0.235

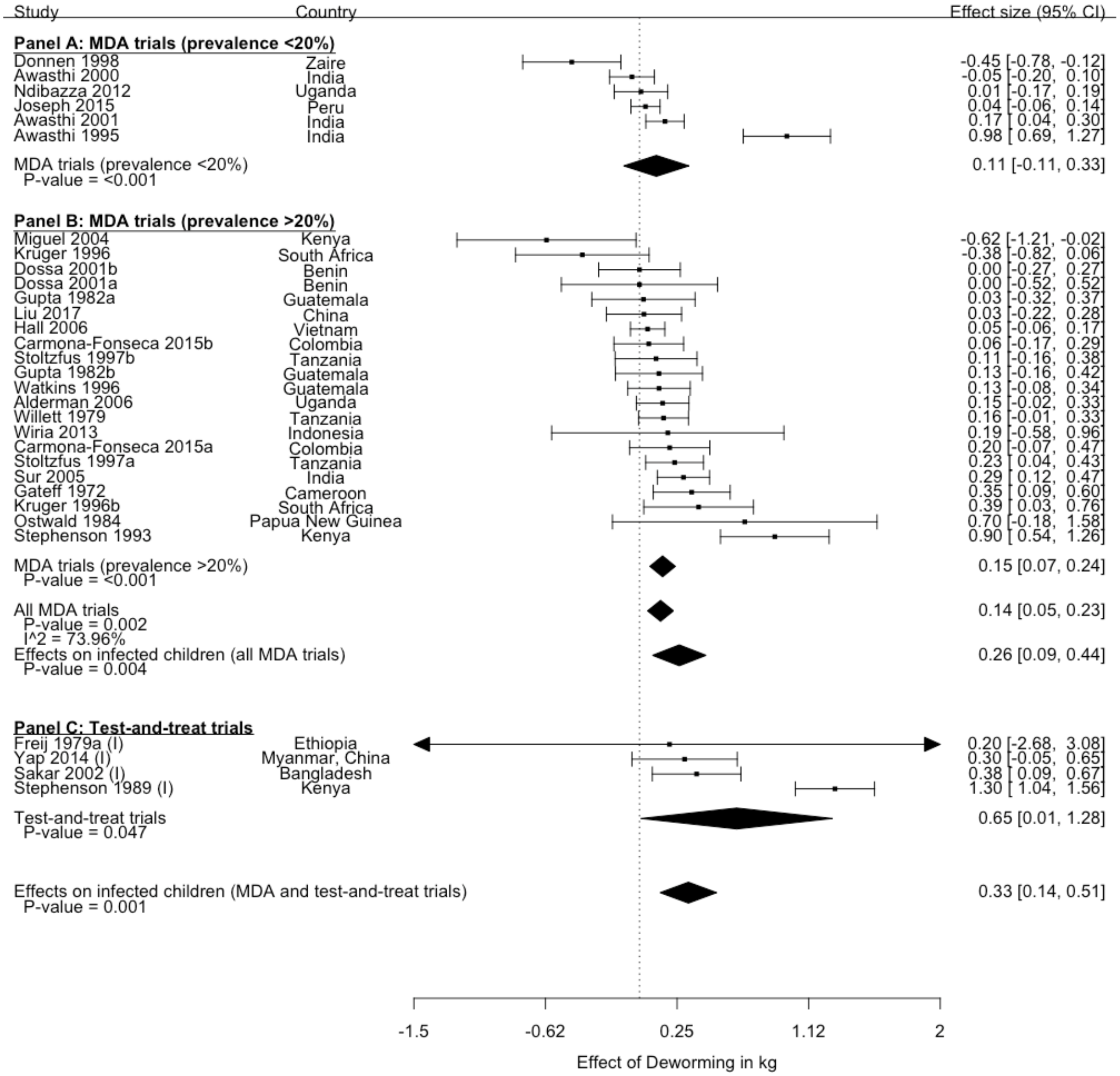
Notes: Estimation method in panels A, B, and C is random-effects. Estimation method in Panel D is random-effects meta-regression, with an indicator variable for MDA as the independent variable. Estimates are based on full sample of MDA trials. Stephenson et al. (1993) is classified as an MDA trial. Point estimates and standard errors from MDA trials have been divided by infection prevalence.

Table 6: Cost-effectiveness analysis

	Deworming MDA (≥20% Prevalence settings)			Deworming MDA (≥50% Prevalence settings)			School feeding			Preschool feeding		
	Avg effect [Avg no. doses]	Avg effect per 2 doses = 2 * ((1) / avg no. doses)	Gain per $1,000 = (2) * (1,000 / cost of 2 treatments)†	Avg effect [Avg no. doses]	Avg effect per 2 doses = 2 * ((4) / avg no. doses)	Gain per $1,000 = (5) * (1,000 / cost of 2 treatments)†	Avg effect [Avg duration, months]	Avg effect per 10 months = 10 * ((7) / avg duration)	Gain per $1,000 = (8) * (1,000 / 41)	Avg effect [Avg duration, months]	Avg effect per 12 months = 12 * ((10) / avg duration)	Gain per $1,000 = (11) * (1,000 / 48.7)
	1	2	3	4	5	6	7	8	9	10	11	12
Weight (kg)	0.154 [3.14]	0.098	144.6 [63.8, 258.7]	0.172 [3.25]	0.106	156.1 [68.9, 279.4]	0.39 [15.3]	0.255	6.2	0.12 [6]	0.240	4.9
MUAC (cm)	0.198 [3.50]	0.113	166.5 [73.5, 298.0]	0.198 [3.50]	0.113	166.5 [73.5, 298.0]	0.31 [23]	0.135	3.3	NA	NA	NA
Height (cm)	0.087 [3.19]	0.054	80.0 [35.3, 143.1]	0.095 [3.36]	0.056	83.0 [36.6, 148.5]	0.38 [15.3]	0.248	6.1	0.27 [6]	0.540	11.1
Hb (g/dl)	0.069 [2.45]	0.056	82.6 [36.5, 147.7]	0.044 [2]	0.044	65.3 [28.8, 116.9]	-0.40 [23]	-0.174	--	0.049 [8.4]	0.07	1.4

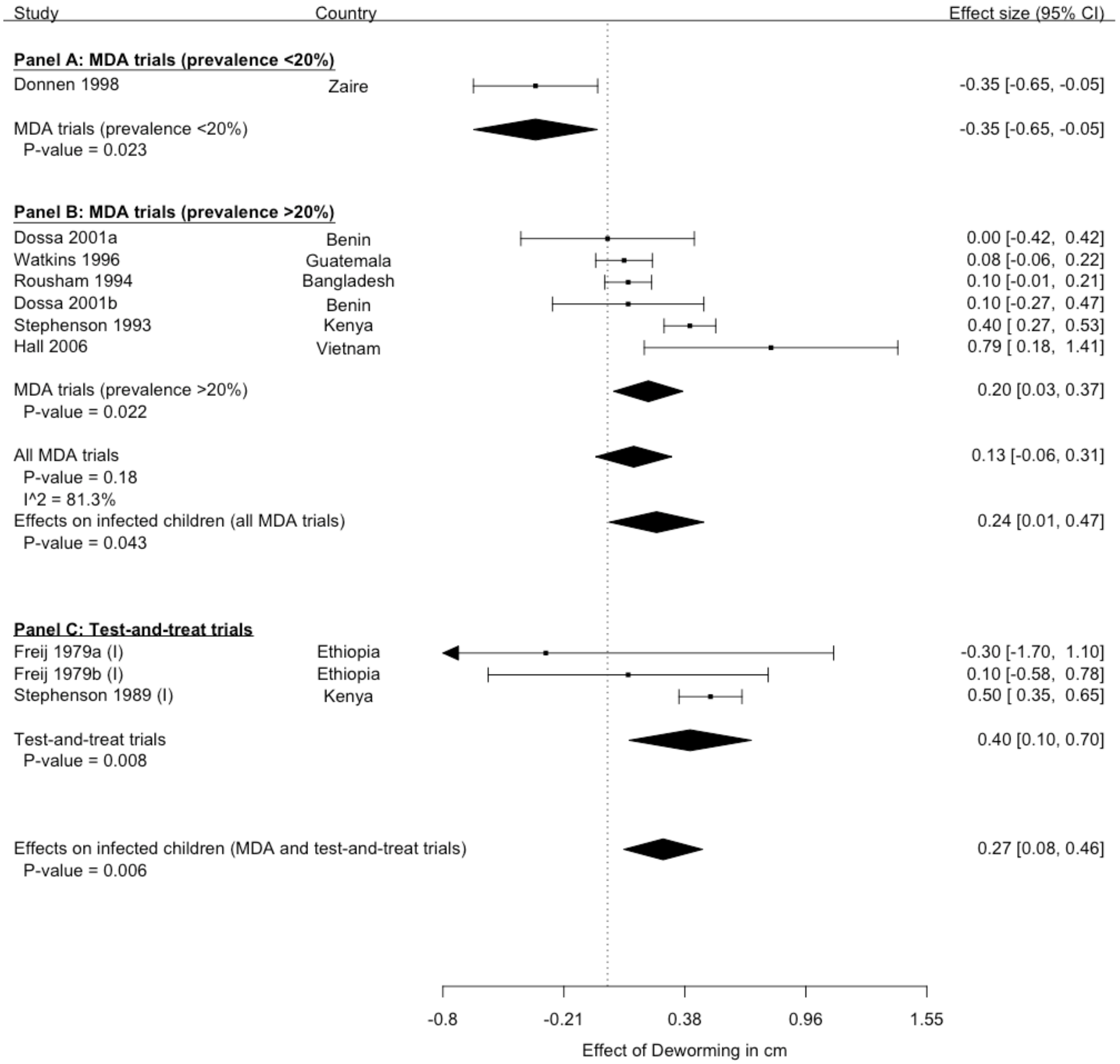
Notes: Estimates of the average child nutrition effects of deworming MDA correspond to our random effects estimates. †We assume a per capita cost of $0.34 for one deworming treatment. This is the current cost estimate for India (GiveWell, 2017), and it incorporates an estimate of the opportunity cost of the time that teachers spend in deworming programs, based on their wages. In square brackets we show a lower and upper bound of the outcome gain per $1,000 spent, using the higher cost per treatment of $0.77 that GiveWell (2017) estimates for African countries (also inclusive of the time of teachers) and the lower cost per treatment of $0.19 in India, if one values the opportunity cost of the time of teachers at one quarter of their wage, respectively. Estimates of the child nutrition effects of school feeding programs in LMICs come from Kristjansson et al. (2007). Estimates for weight and height correspond to random effect estimates. Estimates for MUAC and hemoglobin come from a single study in Kenya (Neumann, 2003). Estimates of the child nutrition effects of preschool-feeding programs in LMICs come from Kristjansson et al. (2015a). Estimates for weight, height, and hemoglobin correspond to random effect estimates, no estimate of the effect on MUAC is provided in the review. $41 is the per capita cost estimate of the daily provision of a ration of 401kcal for a 200-day school year, and $48.7 is the per capita cost estimate of the daily provision of a ration of 397kcal for a calendar year (Kristjansson et al., 2016).

Figure 1: Forest plot of the effect of deworming on weight (kg)



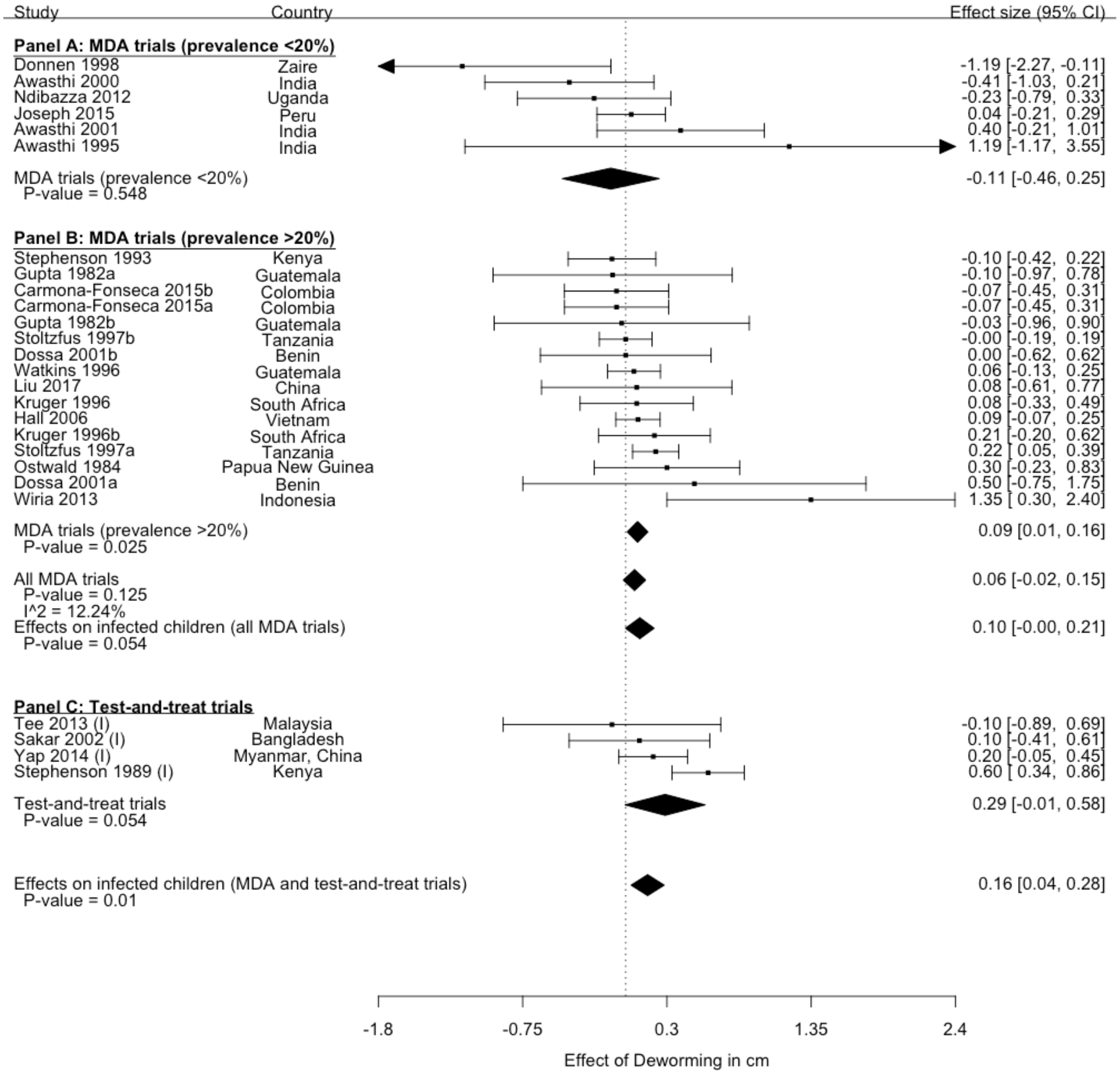
Notes: Panel A shows results from MDA trials conducted in settings where average prevalence is below 20%. Panel B shows results from MDA trials conducted in settings where average prevalence is above 20%. Panel C shows results from test-and-treat trials. We show estimated mean effects for each subgroup, and we also estimate mean effects for all MDA trials (including trials conducted in settings above and below 20% prevalence). In addition, we estimate mean effects for infected children and effects per unit of worm load using all MDA and test-and-treat trials. To estimate the mean effect on infected children, point estimates and standard errors from MDA trials were divided by infection prevalence prior to applying a random effects model. All mean effects are estimated using a random effects model. Arrows indicate that the confidence interval is larger than what is displayed on the graph.

Figure 2: Forest plot of the effect of deworming on MUAC (cm)



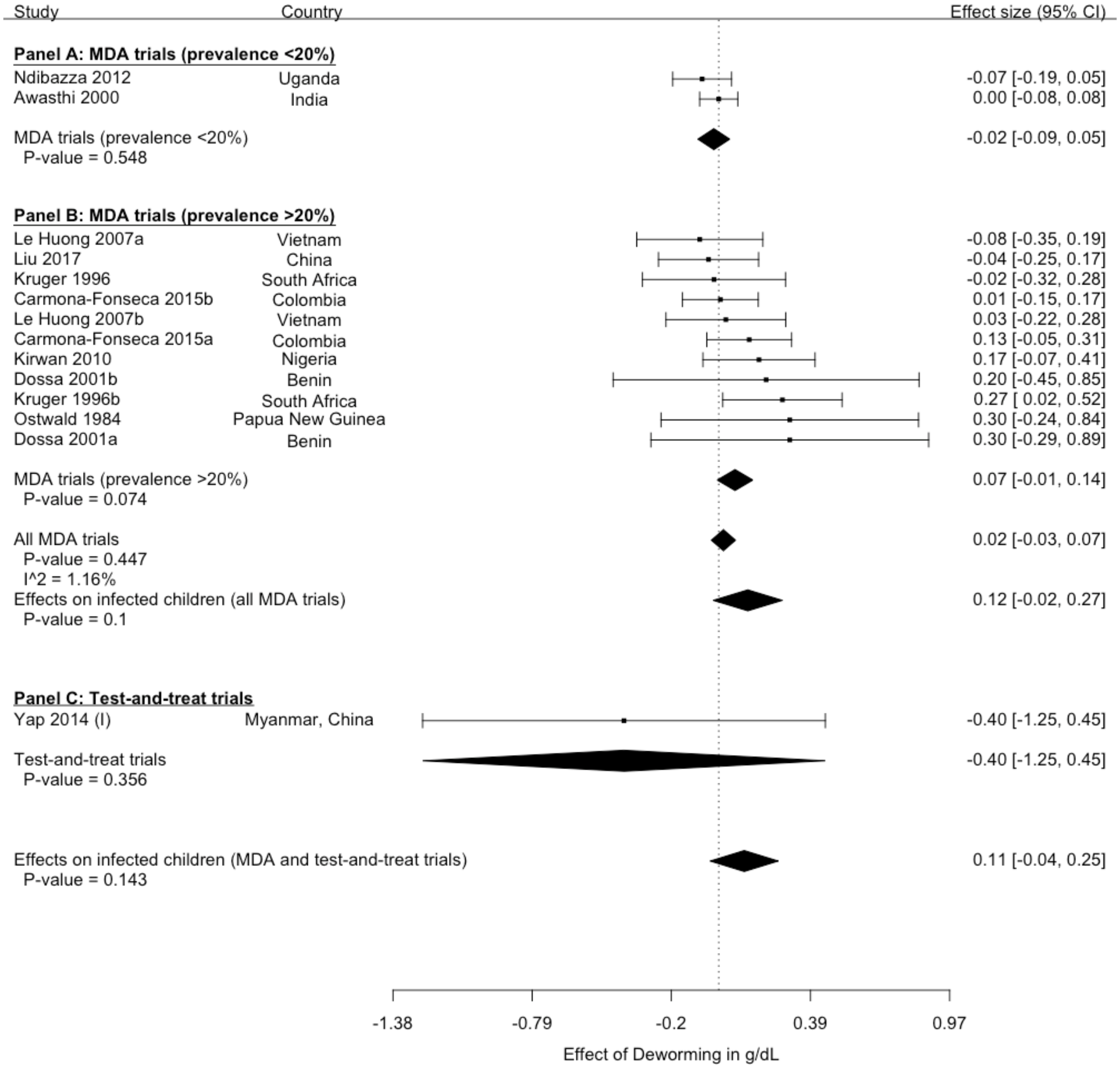
Notes: Panel A shows results from MDA trials conducted in settings where average prevalence is below 20%. Panel B shows results from MDA trials conducted in settings where average prevalence is above 20%. Panel C shows results from test-and-treat trials. We show estimated mean effects for each subgroup, and we also estimate mean effects for all MDA trials (including trials conducted in settings above and below 20% prevalence). In addition, we estimate mean effects for infected children and effects per unit of worm load using all MDA and test-and-treat trials. To estimate the mean effect on infected children, point estimates and standard errors from MDA trials were divided by infection prevalence prior to applying a random effects model. All mean effects are estimated using a random effects model. Arrows indicate that the confidence interval is larger than what is displayed on the graph.

Figure 3: Forest plot of the effect of deworming on height (cm)



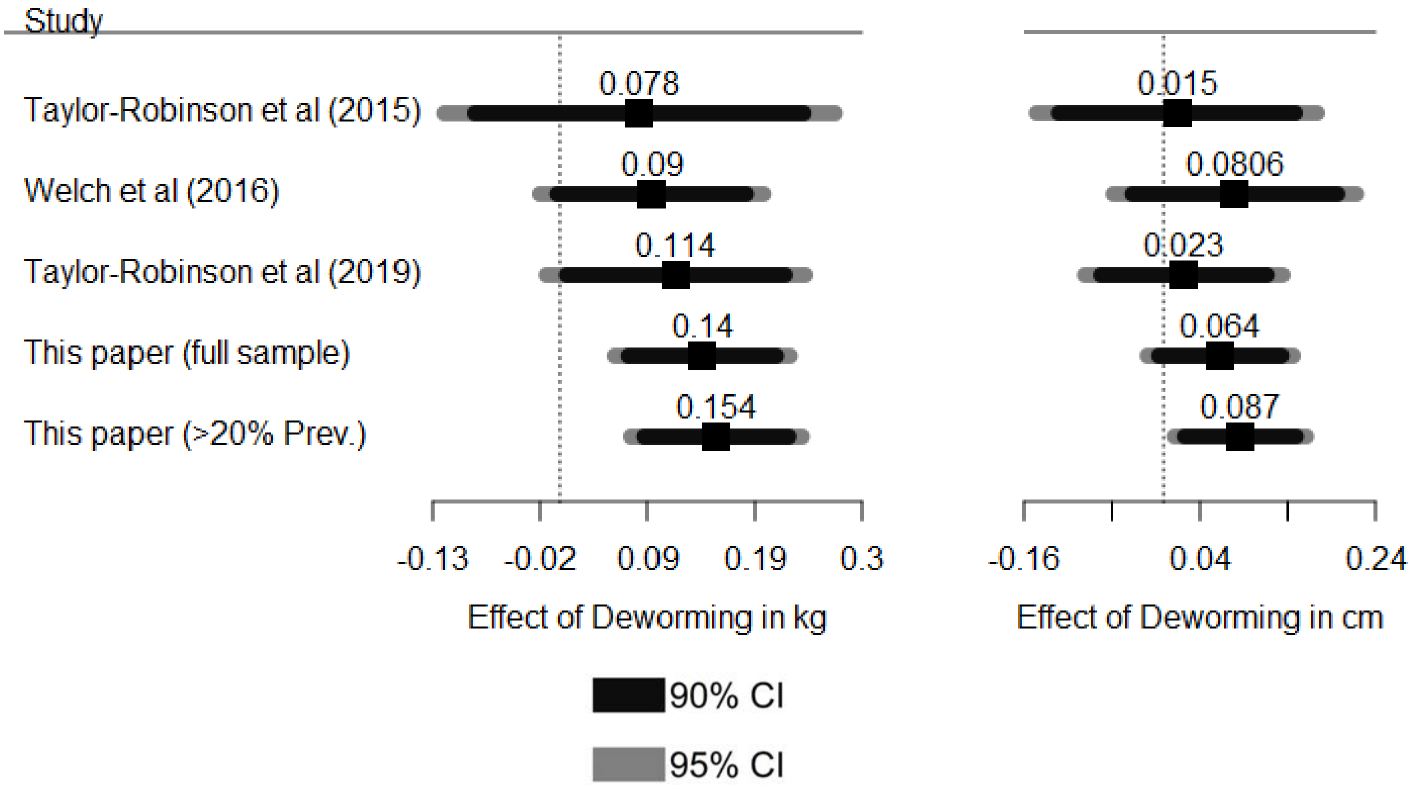
Notes: Panel A shows results from MDA trials conducted in settings where average prevalence is below 20%. Panel B shows results from MDA trials conducted in settings where average prevalence is above 20%. Panel C shows results from test-and-treat trials. We show estimated mean effects for each subgroup, and we also estimate mean effects for all MDA trials (including trials conducted in settings above and below 20% prevalence). In addition, we estimate mean effects for infected children and effects per unit of worm load using all MDA and test-and-treat trials. To estimate the mean effect on infected children, point estimates and standard errors from MDA trials were divided by infection prevalence prior to applying a random effects model. All mean effects are estimated using a random effects model. Arrows indicate that the confidence interval is larger than what is displayed on the graph.

Figure 4: Forest plot of the effect of deworming on hemoglobin (g/dl)



Notes: Panel A shows results from MDA trials conducted in settings where average prevalence is below 20%. Panel B shows results from MDA trials conducted in settings where average prevalence is above 20%. Panel C shows results from test-and-treat trials. We show estimated mean effects for each subgroup, and we also estimate mean effects for all MDA trials (including trials conducted in settings above and below 20% prevalence). In addition, we estimate mean effects for infected children and effects per unit of worm load using all MDA and test-and-treat trials. To estimate the mean effect on infected children, point estimates and standard errors from MDA trials were divided by infection prevalence prior to applying a random effects model. All mean effects are estimated using a random effects model. Arrows indicate that the confidence interval is larger than what is displayed on the graph.

Figure 5: Comparison of the estimated mean impact of deworming MDA across meta-analyses



(a) Weight effect (kg)

(b) Height effect (cm)

Notes: The estimation method in Taylor-Robinson et al. (2015) is random-effects for weight and fixed-effect for height. The estimation method in Welch et al. (2016), Taylor-Robinson et al. (2019), and this paper (for both samples) is random-effects for both outcomes. The main analysis of Welch et al. (2016) is of standardized mean differences, but they present estimates of the mean effect of deworming MDA on weight (in kg) and on height (in cm) in their "Summary of findings table" (p.19), which we use for this graph. We back out the standard errors of these estimates based on the reported confidence intervals. The vertical dotted lines represent an effect size of zero.