Last updated August 2022

Mass Deworming Replicability Adjustment

Contents

Introduction

This document is an extension to SoGive's main analysis of mass deworming. The potentially strong cost-effectiveness of mass deworming relies on evidence from 10-year, 15-year, and 20-year follow-ups to an experiment in Busia, Kenya. A crucial element of our cost-effectiveness analysis is the replicability adjustment, which we apply to the evidence from these follow-up studies. The replicability adjustment answers the question: if we were to run 100 perfect randomised controlled trials (RCTs) (placebo-controlled and no attrition) in exactly the same context as the Busia experiment, what would a hypothetical meta-analysis of these trials conclude? There are several reasons we might expect the conclusions from such a meta-analysis to be less optimistic:

- Since there is only a single experimental context (Busia) that provides direct evidence on the long-run effects of mass deworming, the results could be due to random chance.

    - For example, there may have been a baseline imbalance between the treatment and control groups, although GiveWell senior adviser David Roodman mostly allays such concerns.
    - Ioannidis (2005) argues that the results from a single underpowered but well-performed RCT are only replicated 23% of the time (table 4). However, the average effect size of deworming on earnings and spending in the Busia experiment is large (10.9%), which Ioannidis considers to be an important factor for replication (corollary 2), although the earnings effect is not statistically significant at conventional levels (see "Statistical significance" in our main analysis of deworming for details).
- The long-run results could be explained by a mechanism other than deworming treatment.
    - As GiveWell notes, the promise of deworming tablets, which were distributed in schools, could have encouraged parents of the treatment group to increase their children's school attendance during the experiment. This could potentially have led to a habit of attending school, thereby increasing human capital. The long-run follow-up results would therefore not be explained by the treatment of worm infections.
- It is possible that the results of the follow-up studies have been presented in a way that over-emphasises more positive results, due to publication bias.
- The results are not easily reconcilable with two meta-analyses on the short-run health, educational, and nutritional effects of mass deworming. A Cochrane review finds little evidence of any effect of mass deworming, while a Campbell review concludes that treating schistosomiasis might slightly increase child weight (see "Black box problem" in our main analysis of deworming for details).
- A particular characteristic of the experimental context may have been especially favourable to gains from deworming. For example, one GiveWell staff member suggested that the "marginal product of labor was probably particularly high due to the preceding cohort of prime-age adults experiencing the peak of Kenya's HIV/AIDS pandemic".

The resulting cost-effectiveness of mass deworming after taking into account our replicability adjustment can be found here.

GiveWell's Approach

There are three ways in which GiveWell estimates the likely effect of mass deworming on long-run earnings in a hypothetical meta-analysis of identical trials, outlined here and calculated here.

Method 1 - Mechanisms

Explicitly predicting the effects of particular mechanisms:

- Pros:

    - Precisely includes all the information we have on effect sizes and confidence intervals
    - Provides a strong overview of all the potential mechanisms causing long-run earnings effects
- Cons:
    - Does not consider the potential methodological biases of studies
    - Difficult to take into account the possibility that different mechanisms work together to result in long-run earnings effects

Method 2 - Subjective Bayesian Analysis

Making a series of intuitive *quantitative* predictions based on a holistic view of the evidence:

- Pros:
    - Provides an opportunity to look into related literature, such as effect sizes in similar interventions and replicability rates for studies of similar quality
    - Other more formal methods may lead to misleading results because their assumptions aren't well understood
- Cons:
    - Informal and requires judgement that may be subject to personal bias
    - Difficult to avoid incorporating the same evidence and assumptions into different parts of the analysis (may lead to cross-over)

Method 3 - Subjective Best Guess[1]

Making a best guess based on a holistic view of the evidence:

- Pros:
    - Not limited to providing calculations that are based on strong assumptions
- Cons:
    - Very informal and requires judgement that may be subject to personal bias

SoGive's Approach

We decided against using method 3, which seems to require too much guess work without a structured analysis.

We also decided against method 2, as it still requires a lot of guesswork (though less than method 3), and we worry that multiplying lots of very uncertain numbers together produces a very uncertain final number. Furthermore, it is difficult to know which part of the model should incorporate different forms and quality of evidence. This increases the chance of double-counting particular pieces of evidence or features of the evidence.

[1] See here and here

Method 1 has the advantage of incorporating the evidence in a systematic way, so we use an adapted version of this method. Before launching in, we need to address the issue of "economic losers".

Economic Losers

Before reading this, you may like to read the "Economic Losers" section in our main analysis of deworming. In summary, if dewormed children later gain higher-earning employment, others may miss out on the opportunity to gain higher-earning employment.

To estimate the true effect of mass deworming on productivity and overall welfare, we need to estimate what proportion of the gains from the Busia experiment stem from true productivity gains. A useful proxy might be to calculate what proportion of the total earnings gains stem from self-employment profit, as discussed in the "Economic Losers" section here.

In the 20-year follow up, for the full period between 10 and 20 years after the start of treatment, the average treatment effect on wage earnings was $81, compared to $41 for self-employment profit, while the treatment effect on individual farming profit was zero. The treatment effect on total earnings was $80, which suggests that is *not* simply the sum of its components (wages, self-employment profit, individual farming profit). We might therefore be concerned that the self-employment profit and wage earnings figures are not comparable. However, three factors give us some confidence that these figures are in fact comparable, and that we can calculate self-employment profit gains as a proportion of total earnings gains.

Firstly, the main reason for the discrepancy (i.e. total earnings is not the sum of its components) is that the top 1% of observations are trimmed for each earnings component.[2] This explains why the components do not add up to total earnings, even if they are comparable figures for analytical purposes.

Secondly, the sample sizes were very similar between the self-employment and wage-earnings samples. While the paper reports *average* treatment effects (ATE) for each component, *total* treatment effects (TTE) (i.e. ATE * sample size) capture the overall aggregated effects of deworming. Since the sample sizes were similar, this implies that the ratio of ATEs between the two earnings components is similar to the ratio of TTEs, allowing us to use ATEs for comparison.

Thirdly, not only are the sample sizes similar, but the same trial participants are included in the self-employment and wage-earning samples (p. 9), because "those without any reported earnings in the last year are included in the analysis as zeros". This ensures the two samples are comparable.

We assume that all gains in self-employment profit reflect gains in productivity, since self-employment is not likely to lead to "economic losers". The original ATE on total economic benefit is 10.9%, based on a weighted average of earnings and consumption (spending) gains.

[2] We thank an author of the paper, Edward Miguel, for this insight.

The effect on earnings includes wage earnings, self-employment profit, and individual farming profit.[3] Thus, our new ATE, after accounting for "economic losers", is the original 10.9%, multiplied by the gains in self-employment profit as a proportion of the gains from all components, i.e. 10.9% * $41 / ($41+$81) = 3.66%.[4] An alternative approach would be to calculate gains in self-employment profit as a proportion of *baseline* total earnings. This is a more direct calculation, but ignores the spending results. The result of this alternative approach is 3.31%, similar to the previous approach.

Bayesian Analysis

There are two types of evidence that form our view on the benefits of mass deworming, which are incorporated into a Bayesian analysis, as described under "method 1".

1. The effects of mass deworming on earnings based on the short-run evidence on health, education, and nutrition (Cochrane review and Campbell review meta-analyses).

2. The effects on long-run earnings based on the 10-year, 15-year, and 20-year follow-ups to the Busia experiment.

We particularly value two aspects of the evidence:

1. What is the quality of evidence? Are they randomised controlled trials (RCTs)? Is it a meta-analysis or a single study?

   To offer a point of comparison to other GiveWell recommended charities, we summarise GiveWell's internal and external validity adjustments here. For example, Against Malaria Foundation (AMF) receives a 95% internal validity adjustment from GiveWell. This is based on strong evidence from several RCTs, which are the gold standard for establishing causality in effectiveness research, and some small potential publication bias.

   While the Busia experiment is not technically randomised, GiveWell senior adviser David Roodman demonstrates that the statistical balance of the experiment is robust to "a barrage of tests". Furthermore, he finds that the positive earnings effects are corroborated by a second, truly randomised (smaller) "cost-sharing" experiment in the same setting.

   Finally, Ozier (2018) finds that children born in treated communities during the initial (larger) Busia experiment benefited indirectly through the deworming of their older siblings and neighbours. These children performed better on cognitive tests 10 years later compared to children born in control communities, and the effect was twice as large among children with siblings attending school at the time of the experiment (table 3, p.249). Even if there had been statistical imbalance between treated and control

[3] Note that the effect on farming profit is zero.
[4] The treatment effect on self-employment profit is $41, while the treatment effect on wage earnings is $81.

communities, this would not explain the latter result, since cognition improved most for children whose siblings were dewormed at school, implying that they were less exposed to worms. We therefore conclude that the quality of evidence is relatively strong for the long-run follow-up studies to the Busia experiment.

Meanwhile, all studies included in the meta-analyses on the short-run health, educational, and nutritional effects of mass deworming are RCTs. The key differentiating factor for quality of evidence should therefore be sample size, which we account for by using the confidence intervals from each type of evidence in our Bayesian analysis (see section “Incorporating Direct Long-Run Effects”).

2. How direct is the evidence? In other words, how much does this tell us about the effects of mass deworming on earnings without having to look into other evidence?

   The evidence for gains in long-run earnings/spending are more direct in the Busia experiment’s follow-up studies compared to the Cochrane and Campbell reviews, which focus on short-run health, educational, and nutritional effects. This could motivate placing additional weight on the results from the Busia experiment follow-ups, as it gives us more precise information about the potential long-run effects of mass deworming.

   However, considering the uncertainty around “economic losers” described above, we choose not to place any extra weight on the Busia experiment follow-ups. The health, educational, and nutritional effects recorded in the Cochrane and Campbell reviews are not likely to be subject to “economic losers”, since any improvements in these short-run measures, such as weight gain, should generate true gains to productivity. Therefore, any improvements in short-run health, education, or nutrition should lead to societal gains in long-run earnings and spending. In contrast, due to economic losers, the earnings gains from the long-run follow-ups to the Busia experiment may exceed the productivity gains that are generated by short-run impacts on health, education, or nutrition. This indicates that additional earnings gains beyond those short-run benefits may be subject to “economic losers”.

New Replicability Adjustment

To calculate our replicability adjustment, we incorporate the two types of evidence described above.

Short-run Health, Educational and Nutritional effects

GiveWell’s analysis of potential mechanisms provides a summary of three possible mechanisms driving the earnings and consumption (spending) effects from the Busia experiment.

Weight Gain

Firstly, the most robust evidence of the short-term effects of multiple-dose mass treatment for soil-transmitted helminthiasis (STH)[5] is an increase in child weight, a measure of nutrition, as described by GiveWell senior adviser David Roodman. A 2015 Cochrane review meta-analysis finds a weight gain of 0.08 kg (95% confidence interval (CI) -0.11 to 0.27) across 10 trials, while a 2016 Campbell review finds a similar treatment effect of 0.09 kg (95% CI -0.09 to 0.28). However, a third meta-analysis, Croke et al. (2016) argues that more data should be included, resulting in a treatment effect of 0.134 kg (95% CI 0.03 to 0.24). Roodman reconciles these somewhat contradictory results by concluding they are mostly consistent at around a 0.1 kg weight gain. More recently, an updated 2019 Cochrane review (see "Summary of findings for the main comparison") similarly concludes from 18 multiple-dose trials that the weight gain from deworming for STH is 0.11 kg (95% CI -0.01 to 0.24), though the quality of evidence is considered "very low".

Recall this exercise is assessing the likelihood that mass deworming has positive effects on productivity in a setting similar to the Busia experiment, where the worm burden was particularly high. Thus, we need to scale up the ~0.1 kg weight gain from the meta-analyses to a higher worm burden setting.[6] The worm burden in Kenya during the Busia experiment was around doubly as severe as other trials included in the 2019 Cochrane Review, as measured by infection prevalence of STH.[7] This implies that the weight gain might be doubly as large in a setting similar to the Busia experiment. The 2019 Cochrane Review (see "Analysis 2.1") corroborates this theory, finding that multiple-dose mass deworming in *high burden* settings raises child weight by 0.2 kg (95% CI -0.17 to 0.56), up from 0.1 kg.[8] We therefore estimate a weight gain effect of 0.2 kg in a setting similar to the Busia experiment.

How convincing is the argument that mass deworming generates these weight gains, given that the effect is not statistically significant at conventional levels in either the 2019 Cochrane or Campbell review? We might be concerned that the statistically *non-significant* positive effects on weight are down to random chance, and therefore do not indicate impact. However, it is worth noting these estimates are based on mass deworming across a whole population, including some children that are not infected. This is most reflective of how GiveWell-recommended charities carry out their work. Nonetheless, it may be informative to examine the effects of treating children *known to be infected*, since this gives us an insight into whether nutritional gains for some children are at all plausible.

---

[5] GiveWell-recommended deworming charities also treat children for schistosomiasis, a waterborne infection caused by parasitic worms. Most of the evidence on mass deworming pertains to treatment for STH.

[6] We will later scale back down the cost-effectiveness of mass deworming based on the intensity of worm burdens in areas where GiveWell-recommended charities currently operate.

[7] Ideally we would use prevalence of moderate-to-heavy infections, which are a stronger predictor (see "Lighter worm burdens") of harms from worm infections. But we were only able to obtain this data for two studies (Miguel and Kremer 2004, Joseph 2015 (Table 3)).

[8] The Busia experiment (Miguel and Kremer 2004) had a similar average infection prevalence to other studies in high burden settings.

Targeted deworming for children *known to be infected* with STH increases weight by 0.49 kg (95% CI 0.07 to 0.90) according to the Campbell Review ("Results" section), or 0.75 kg (95% CI 0.24 to 1.26) according to the 2015 Cochrane review ("Data and Analyses" section - "comparison 1"). Both are statistically significant at the 5% level. The Campbell review also finds a statistically significant 1.47 kg gain (95% CI 0.82 to 2.11) for treating schistosomiasis, which GiveWell-recommended charities treat in addition to STH in some of their programmes.[9] We use these results to approximate the effect size on children with moderate-to-heavy worm burdens, since children known to be infected are less likely to have light infections, which are often asymptomatic (p.161). To apply these results to the Busia experiment, we use GiveWell's finding that 12.77% of children in the control group had moderate-to-heavy infections during the experiment. Therefore, in a Busia-like setting, we would expect 12.77% of children to be known to be infected. This implies that we should scale down the results from the Campbell and Cochrane reviews (for children known to be infected) to apply them to a Busia-like setting.[10] This produces an effect size of 0.063 kg (Campbell) or 0.096 kg (Cochrane) for STH, and 0.19 kg for schistosomiasis. These results could plausibly be reconciled with the previous approach, which indicated a weight gain of around 0.2 kg in a high burden setting.

Since the second approach was more to assess plausibility than to obtain a precise estimate, we now return to an estimate of a 0.2 kg weight gain for mass deworming in a *high burden* setting, based on the 2019 Cochrane review. We also need to convert 0.2 kg gain in weight into standard deviations, because 0.2 kg means a lot more for a young child than an older child. To make this conversion, we use the Campbell review results described in Roodman's analysis[11]. The Campbell review finds a 0.09 kg gain, corresponding with a 0.05 standard deviation (sd) gain in weight-for-age (i.e. a 0.05 gain in the weight-for-age z-score). This implies that 0.2 kg is approximately equivalent to 0.2 / 0.09 * 0.05 = 0.11 sd. We estimate a 90% confidence interval of 0.06 to 0.16 sd (in a Busia-like high burden setting). We keep the same width CI as the results from the Campbell review (see "Nutrition"), because it seems most plausible that scaling up the effect size (due to the high worm burden setting) should not substantially change the precision of our estimate.

GiveWell estimates that a 1 sd increase in child weight-for-age (i.e. a one unit increase in the weight-for-age z-score) is associated with a 3% increase in earnings (90% CI 0.1 to 11, lognormal distribution). This gives us an effect size on earnings of 0.11 sd * 3% = 0.33%, under the assumption that the effect of weight gain scales linearly.[12] Our version of GiveWell's simulation model produces a 0.33% increase in earnings with a standard deviation of ~0.8% (box PII11).[13]

---

[9] For example, see GiveWell's report on Deworm the World Initiative.
[10] For example, the Campbell review results for STH would produce an effect size of 0.49 kg / (100%/12.77%) = 0.063 kg.
[11] The initial published version of this meta-analysis reported the effect size in standard deviations, but the current version of the Campbell review reports the effect size in kilograms.
[12] i.e. we assume that a 0.11 sd weight gain is 11% as beneficial as a 1 sd weight gain.
[13] Note that the simulation produces a different result each time, so this is an approximation.

Suppose that the Busia experiment follow-up results did not exist, and our opinion of mass deworming was based only on the weight gain evidence. This would indicate a 0.33% increase in long-run earnings as a result of childhood deworming over "a period of about 12 months" (p.1). To obtain a replicability adjustment for the context of the Busia experiment, in which the treatment group received an additional 2.41 years of deworming (p.8), we multiply the 0.33% increase in earnings by 2.41 years of treatment. However, as GiveWell notes, additional years of treatment may have diminishing returns, since full re-infection typically takes longer than the time between treatments.[14] In the studies of weight gain contained in the Campbell review, treatment group children were treated for 12 months, compared to zero for the control group. Conversely, in the context of the Busia experiment, children in the treatment group received deworming for 4.1 years compared to 1.7 years in the control group.[15] We therefore apply GiveWell's model of diminishing returns to additional years of treatment, which gives us an adjustment factor of 79.9%. This implies that treatment for the duration of the Busia experiment should have 2.41 * 79.9% = 1.9x the impact of treating children for one year.

Overall, based on the evidence on weight gain alone, this implies that 2.41 years of treatment produces an earnings effect size of 0.33% * 2.41 * 79.9% = 0.64%. This gives us a replicability adjustment of 0.64% / 10.9% = 5.8%, where 10.9% is the original effect size from the 20-year follow-up to the Busia experiment. Applying this result to our cost-effectiveness analysis, the "Weight Gain Only Model" shows that mass deworming charities would still be competitive with SoGive's gold benchmark of £50 to double someone's consumption (spending) for one year, ranging from £42 to £74.

Incorporating Direct Long-Run Effects (Follow-ups to the Busia Experiment)

Weight Gain + Long-Run Earnings Results

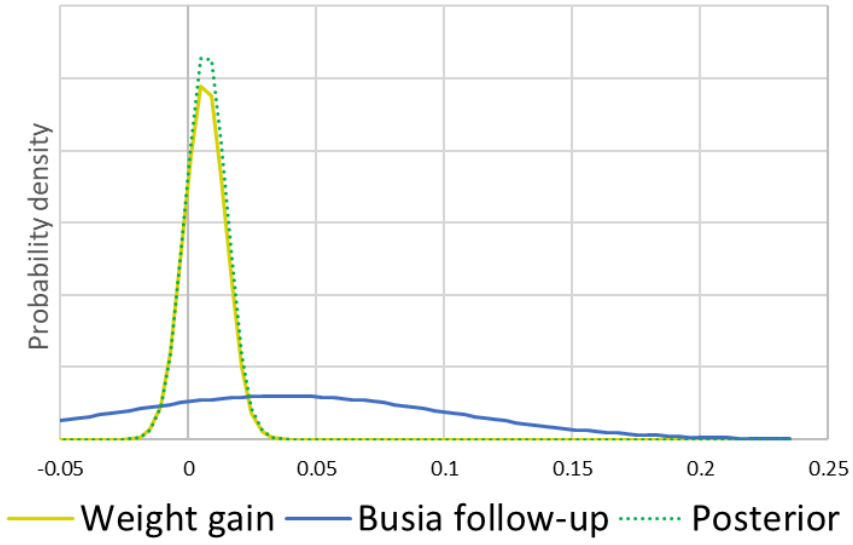
We now incorporate the results from the 20-year follow-up to the Busia experiment, using GiveWell's "Bayesian decider tool". For the first piece of evidence, we use the evidence on the effects of mass deworming on weight gain, as described above. This gives us an effect size of 0.64% on long-run earnings/spending, with a standard deviation of 0.8%. For the second piece of evidence, the effect size on long-run earnings/spending from the Busia experiment is 10.9%, with a standard deviation of 6.61%. However, after accounting for "economic losers", we estimate that the effect size is 3.66% (see section "Economic Losers"). For this Bayesian analysis, we use an effect size of 3.66% and a standard deviation of 6.61%. This implies a wide confidence interval with a large part of the distribution in the negatives, which is probably

[14] At 12 months post-treatment, Jia et al. 2012 report that infection prevalence is 94%, 82%, and 57% of pre-treatment prevalence for A. lumbricoides, T. trichiura, and hookworm (species of STH), respectively. Children typically receive treatment for STH at least once per year, in line with WHO recommendations (p.41). Therefore, at 12 months post-treatment, we would expect infection prevalence to remain below pre-treatment levels, reducing the effect of deworming when children are treated for more than one year.
[15] After accounting for coverage rates in the Busia experiment's treatment and control groups, children were dewormed for 2.86 years instead of 1.18 years. For studies of weight gain in the Campbell review, in the absence of data on treatment coverage, we assume children in the treatment group were treated for one full year, compared to zero months of coverage in the control group.

unrealistic[16], but captures the uncertainty around using just a single study (compared to the Campbell review meta-analysis used for the effects of mass deworming on weight gain). The higher standard deviation (6.61% compared to 0.8%) therefore accounts for our relative lack of confidence about the true effect size. Since the "Bayesian decider" allocates more weight to evidence with a lower standard deviation, we judge that any reduction in the 6.61% figure would assign too much weight to the long-run follow-up results from the Busia experiment.

As shown in Figure 1, the new evidence from the Busia experiment (3.66% earnings effect) provides very little update to our estimate from the weight gain evidence (0.64% earnings effect), which is more precisely estimated (i.e. lower standard deviation). The posterior effect size of mass deworming on earnings is now 0.68% (standard deviation 0.79%) for 2.41 years of deworming. This implies a replicability adjustment of 0.68% / 10.9% = 6.24%, i.e. we expect a perfect meta-analysis of RCTs in the context of the Busia experiment to produce an effect size 6.24% as large as the original Busia results. This replicability adjustment is slightly higher than in the "Weight Gain Only Model" (5.8%), in which the evidence from the Busia experiment is eliminated entirely. This update therefore makes mass deworming charities look slightly more cost-effective, reducing the cost per year of doubled consumption (spending) from £42-74 to £39-67.

Figure 1



As mentioned earlier, GiveWell also explores two other mechanisms - years of schooling and IQ. Bundy 2013 (p.145-6) argues there may be subtle cognitive benefits to deworming, which may not be reflected in short-term measures such as weight gain and school attendance. Another reason for exploring other mechanisms is that no weight gain was recorded for children treated in the Busia experiment. However, Roodman finds that initially *underweight* treated

[16] Indeed, David Roodman argues convincingly in support of the positive follow-up results to this experiment, by demonstrating that the experiment provided the most benefit at elevations where worm burdens were heaviest.

children gained weight relative to the control group (while treated children initially *not* underweight lost weight relative to the control group). Long-run earnings gains were also concentrated among these initially underweight children. This suggests that weight gain may have played a part in the long-run earnings gains, though Roodman warns against interpreting this as a causal relationship.

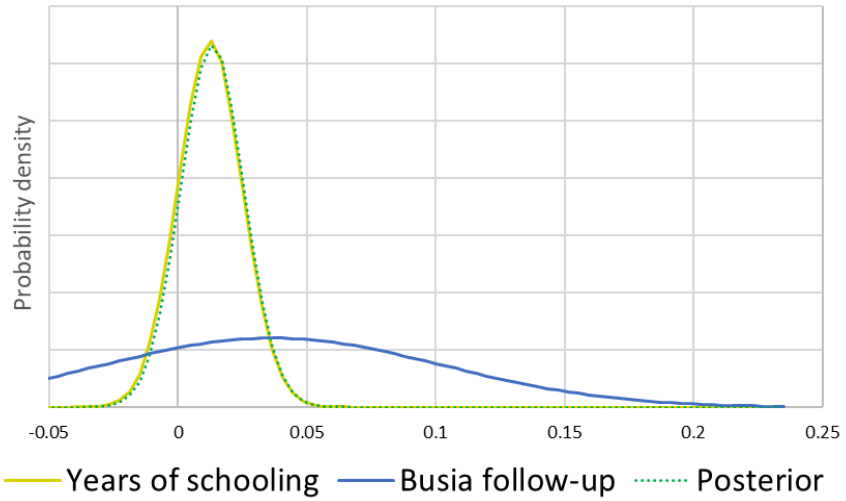
Years of Schooling

The 10-year follow-up to the Busia experiment also finds a 0.15 increase in years of schooling (90% CI -0.085 to 0.385).

In SoGive's review of the literature on returns to education, we estimate that an additional year of education leads to approximately an 8.3% increase in annual earnings in a low- and middle-income country context. We subjectively estimate a 90% confidence interval of 3.3% to 13.3%.

Overall, this implies a 0.15 years * 8.3% = 1.25% earnings effect of mass deworming via the "years of schooling" mechanism. The simulation tool produces a standard deviation of ~1.25% (see box PII8). Since this estimated effect on years of schooling comes from the Busia experiment, it reflects 2.41 years of treatment. For the second piece of evidence in the Bayesian decider, as above (see "Weight Gain + Long-Run Earnings Results"), we use a mean effect size of 3.66% (standard deviation 6.61%) based on the long-run earnings results from the Busia experiment. The Bayesian decider returns an posterior estimate of 1.33% (standard deviation 1.23%) for 2.41 years of treatment, as shown in Figure 2. This implies a replicability adjustment of 1.33% / 10.9% = 12.20%.

Figure 2

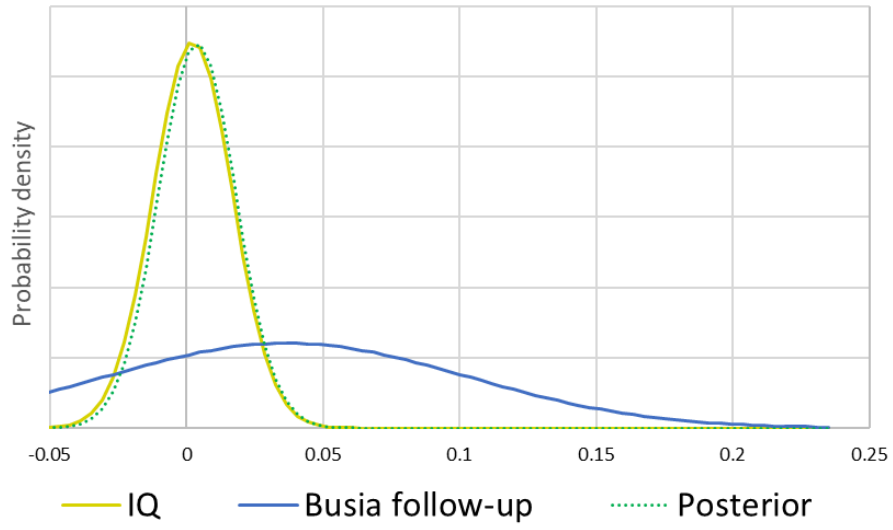


IQ/Cognition

Lastly, GiveWell reviews three studies that report the effect of mass deworming on IQ/cognition. They take a weighted average of these studies based on their similarity of worm burden to the Busia experiment, which amounts to a ~0.2 point increase in IQ (90% CI -2.9 to 3.3). Our work on the effects of IQ on earnings indicates that a 1 point increase in IQ is associated with a 0.70% increase in earnings.[17] We subjectively estimate a 90% confidence interval of 0.2% to 1.2%. Based on these estimates, a 0.2 point increase in IQ is associated with a 0.2 * 0.70% = 0.14% increase in earnings. The simulation model produces a standard deviation of ~1.45% (box PII12).

In the three studies used to measure the effects of mass deworming on IQ, treated children were treated over the course of around 14 months (1.17 years) on average. To estimate the effect of 2.41 years of treatment, we first multiply 0.14% by 2.41/1.17. However, we also need to account for diminishing returns to additional years of treatment, as noted above (see "Weight Gain"). We therefore apply GiveWell's model to account for these diminishing returns, which gives us an adjustment factor of 80.4%.[18] This produces an effect size of 0.14% * 2.41/1.17 * 80.4% = 0.23%. For the second piece of evidence in the Bayesian decider, as above (see "Weight Gain + Long-Run Earnings Results"), we use a mean effect size of 3.66% (standard deviation 6.61%) based on the long-run earnings results from the Busia experiment. Merging these two forms of evidence, the Bayesian decider returns a posterior estimate of 0.39% (standard deviation 1.42%), which implies a replicability adjustment of 0.39% / 10.9% = 3.58%.

Figure 3



[17] Each standard deviation (SD) of IQ is equivalent to 15 IQ points. Therefore, a 10.45% income effect per SD change in test score (IQ) is equivalent to a 0.70% income effect per IQ point.
[18] Note that GiveWell's model calculates the infection rate after 1.18 years, rather than after 1.17 years, but these are close enough for this approximate exercise.

Other Mechanisms

There are other plausible mechanisms that could have led to the long-run earnings results from the Busia experiment. One possibility is that worm infections as children lead to health problems as adults, hindering future productivity. However, we are not aware of any evidence that poor health due to childhood worm infections persists into adulthood.

We also considered long-run literacy and numeracy impacts from Croke 2019. The study shows no treatment effect across the whole sample, although the effect size was quite large in the female sub-sample. We omitted this study, however, due to concerns around potential migration across communities, which could have led some treated children to be placed in the control group and some control children in the treatment group. This would likely place downward bias (p.2166) on the results due to attenuation bias from classical measurement error. Furthermore, the worm burden was lower than in the Busia setting (see p.15), which makes the results more difficult to interpret with regard to its implications for the Busia experiment.

Conclusion

Merging the Results

The final question is how to merge these three mechanisms (weight gain, years of schooling, cognition) into one replicability adjustment. GiveWell approached the task by taking a weighted average based on subjective weights. We use a more formulaic approach to avoid using discretion. We also take a weighted average, but we weight each mechanism using an "inverse-variance weighted average". Each mechanism is weighted by the inverse of its posterior variance (i.e. square of the posterior standard deviation), which accounts for the level of precision we have about that estimate. This is a conservative approach. Since some treated children may benefit from more than just one mechanism, it may be justifiable to add together the effects from different mechanisms. For example, a child might benefit from weight gain through greater size and physical robustness, which could lead to improvements in manual labour output, while also benefiting from improved cognitive skills. However, since the three mechanisms are not likely to be completely independent[19], we take this more conservative approach (inverse-variance weighted average).

Outcome

The outcome of using this "inverse-variance weighted average" approach is that the weight gain mechanism receives most of the weighting, because it has the most precise posterior estimate. This produces an overall replicability adjustment of 7.2%. In other words, based on the evidence

[19] For example, additional years of schooling may be a symptom of improved cognition, without improving long-run earnings in itself. Or alternatively, nutritional gains (reflected by weight gains) could lead to improvements in both cognitive and non-cognitive skills; but if non-cognitive skills are more important in the labour market, cognitive skills (IQ) may not have a causal impact on long-run earnings.

on nutrition, cognition, and schooling mechanisms, and updating these priors based on the 20-year follow-up results to the Busia experiment, we would expect the effect of 2.41 years of deworming on earnings to be 7.4% of the effect size from the 20-year follow-up. For comparison, when using this same mechanisms approach, GiveWell obtains a replicability adjustment of 15%. They also give some weight to more informal approaches, which gives them a final replicability adjustment of 13%.

How does this replicability adjustment model differ from GiveWell's?

Details can be found throughout this report, and in this sheet.

- We only used the "Mechanisms method" from GiveWell's model, described above (see "GiveWell's Approach").
- We accounted for "economic losers" for our estimate of long-run earnings effects of mass deworming from the 20-year follow-up to the Busia experiment.
- We recalculated the effect on earnings from the weight gain mechanism using David Roodman's analysis and using estimates of weight gain effects in a high worm burden context.
- We updated the effect of an additional year of schooling on earnings, based on a SoGive review of the literature.
- We updated the effect of increasing IQ on earnings, based on a previous SoGive review of the literature.
- For the weight gain and IQ mechanisms, we apply downward adjustments to account for diminishing returns to additional years of treatment.
- We used the "inverse-variance weighted average" approach to weigh each mechanism against each other.

Which parameters explain most of the divergence between our replicability adjustment and GiveWell's?

- We first follow GiveWell's mechanisms model, which produces a replicability adjustment of 12.7%.[20]
- We then account for "economic losers", which reduces the replicability adjustment further from 12.7% to 8.6%. This is because the follow-up studies to the Busia experiment provide a smaller update to our estimated earnings/spending effect size if we believe that the true effect on long-run productivity is smaller.
- We adjust upwards the effect on long-run earnings based on the weight gain mechanism, accounting for the fact that the Busia experiment was conducted in a high

[20] GiveWell's published figure is 15.0%, which appears to be an error. GiveWell's mechanisms model suggests that the effect of one year of deworming on earnings through the weight gain / nutrition mechanism is ~0.15% (cell PII5). Therefore, the effect of 2.41 years of deworming should be 0.15%*2.41=0.36%, rather than the published figure of 1.2% (cell F2). This produces a replicability adjustment of 4.5% for the weight gain / nutrition mechanism, and 12.7% overall for the mechanisms approach.

worm burden environment. This increases the overall replicability adjustment for the mechanisms method from 8.6% to 9.8%.

- We also update the effects of IQ and years of schooling on earnings based on SoGive literature reviews. These changes reduce the overall replicability adjustment from 9.8% to 9.2%.
- We adjust downwards the effects from the weight gain and IQ mechanisms to account for diminishing returns to additional years of treatment. This reduces the overall replicability adjustment from 9.2% to 8.6%.
- We use the "inverse-variance weighted average" approach in place of GiveWell's subjective weights on each mechanism. This reduces the replicability adjustment from 8.6% to 7.2%. Since the results are most precisely estimated (narrower confidence intervals) for the nutrition (weight gain) mechanism, we believe that this mechanism provides the most accurate estimate for long-run productivity effects, and is therefore given additional weight. We also give zero weight to an unidentified "other" mechanism, to stay closer to the empirical evidence.

Overall, our best guess is that the effect of mass deworming on earnings/spending is 10.9% * 7.2% = 0.8%, compared to GiveWell's best guess of 10.9% * 13% = 1.4%. This makes us slightly less optimistic about the cost-effectiveness of mass deworming, ultimately reducing its cost-effectiveness by around one-third.

Caveat

If this model was badly wrong, what would be the most likely culprit? The final result is very sensitive to our estimate for the effect of a 0.2 kg weight gain (in a high worm burden setting) on long-run earnings, because the other mechanisms (IQ and years of schooling) have wider confidence intervals. The estimated effect on earnings is based on limited evidence from early life growth intervention literature reviewed by GiveWell, which leads to very uncertain outputs. If this weight gain does not generate the productivity gains we expect (e.g. perhaps in the long-run, treatment leads to greater body size, but no substantial improvements to cognition, health, or physical ability), mass deworming may in fact provide little overall economic benefit.