Deworming replicability adjustment (JS)

Links to other documents:

- Google sheet with rough calculations of 90% confidence intervals depending on mean and standard errors of the estimates.
- Guesstimate sheet with monte carlo simulations for the calculations below

Summary

Baird 2016, a long run follow up to Miguel and Kremer, finds a large direct effect of mass drug administration deworming on long run income. However, the evidence for any mechanisms appears to be weak. How do we reconcile these?

In this document, I look at the evidence for MDA having an effect along three of the most plausible causal pathways:
(i) improved cognition
(ii) increases in weight-for-age
(iii) increases in schooling (grade completion)

I attempt a Bayesian model (i.e. using confidence and credible intervals) of the distribution of possible effect size through each of these pathways. I then update based on the effect sizes from the pooled KLPS2 and KLPS3 survey results on income.

I conclude that:

- **Weight gain:** It seems unlikely that increases in weight play more than a negligible role in increases in future income. The results from both Miguel and Kremer 2004, and the Campbell meta-analysis aren't large enough to explain anything but a very small effect on earnings. One assumption here is that Stephan's model of the link between weight-for-age and income is broadly correct, but I've put fairly wide confidence intervals on it and it still seems unlikely to be able to explain the effects.
- **Cognition:** Improvements in tests of cognition may explain some of the effect. The evidence that deworming improves cognition finds a null result, but with fairly wide confidence intervals. This makes it plausible that deworming increases future income through improved cognition, without affecting weight gain. While the children dewormed in Busia were above the age at which I'd guess deworming could substantially affect brain development, the results may be driven by younger children living in the treatment area. Taking the probability distributions at face value implies a **14% replicability adjustment**. I take this as a lower bound on my replicability adjustment, because it involves the conservative assumption that improving short term cognitive scores is the *only* mechanism through which deworming could affect long run income.
- **Schooling.** Baird 2016 is the only study which attempts to estimate effects on years of schooling completed. It finds positive, but insignificant results, with a wide confidence interval. If we assume improvements in schooling was the mechanism, it implies a **16% replicability adjustment**.
- I have not attempted to model whether haemoglobin is a plausible causal pathway because my impression from Josh is the evidence is weak. I may return to this in future.[1]

[1] I got started on it but deprioritized

- I place low weight on the possibility that there is another biological mechanism which doesn't correlate with *any* of the short term indicators studied: weight, height, cognition, or hemoglobin. I believe the burden of proof is on those hypothesizing these effects to tell a story about how that might happen. (Alexander disagrees somewhat, and cites as evidence that there have been lots of times drugs have reduced mortality without us knowing anything about the causal mechanism -- I'd need to think about this more).

My overall replicability adjustment is 16%. I chose the highest value from the three plausible causal pathways because only one causal pathway is required.

UPDATE (11 July 18):

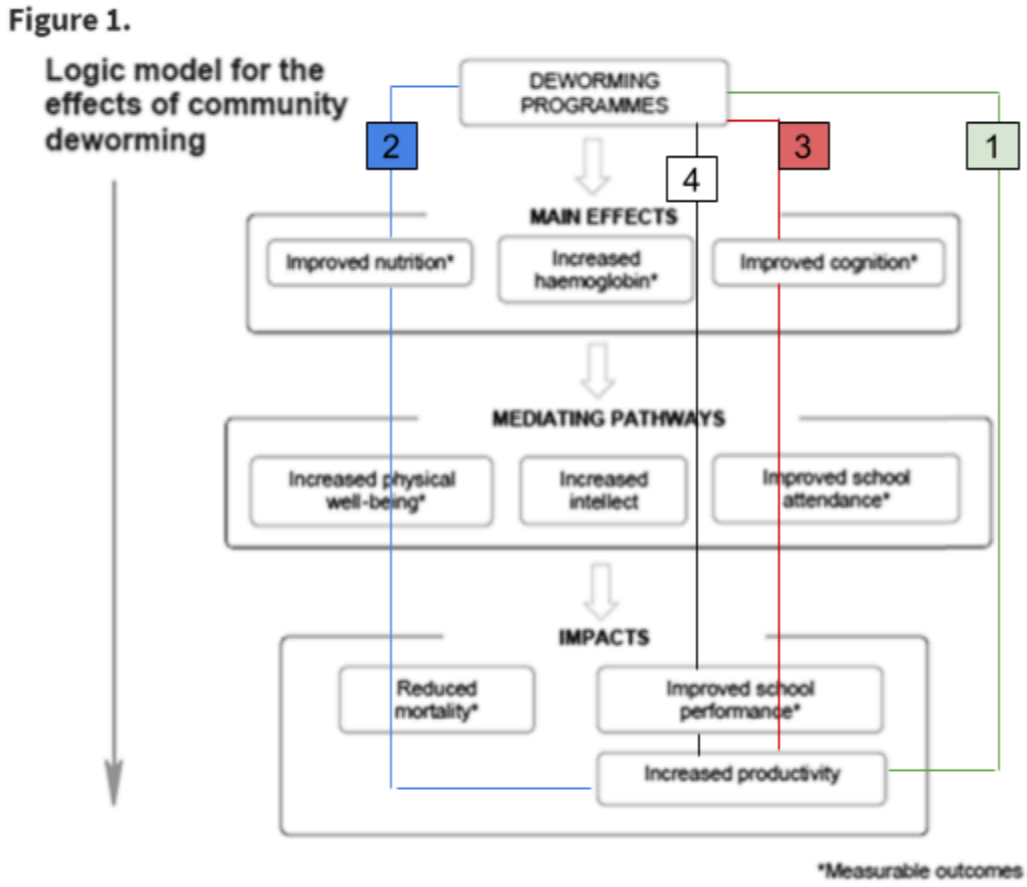
I started doing this analysis before thinking hard about how we should update based on KLPS4 consumption data.

Four different data points

I review four different pieces of evidence we have:

(1) The direct evidence of the effect of MDA on long run earnings.
(2) Potential increases in nutrition indicators
(3) Potential increases in performance on cognition tests
(4) Potential increases in years of schooling (grade completion).

Figure 1.



(*From JSG, 2017*)

The direct evidence of the effect of MDA on long run earnings.

The pooled results of KLPS2 and KLPS3 give a central estimate of a 0.143 increase in ln(income).[2] I calculated the 90% confidence interval as between 0.045 and 0.233. **I.e. between a 4.5% and 23% effect on income**.

Nutrition

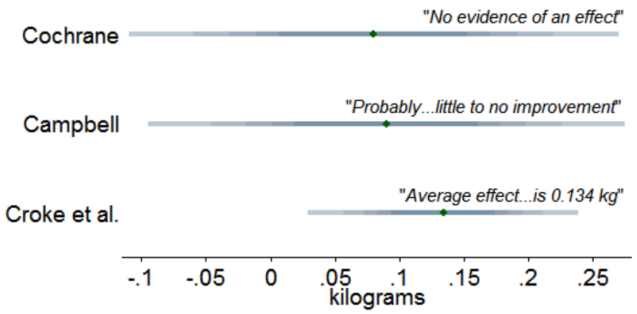
There are two pieces of evidence we might look at to establish how likely it is that increased weight is a plausible mechanism for the increases in long run earnings in Baird 2016.

The first is the results from various meta-analyses of the effect of deworming on weight. The second is the results from Miguel and Kremer itself. It seems implausible to me that weight gain is a mechanism unless it showed up in the same context as the KLPS studies (or at least there were wide confidence intervals).

[2] For further details on the pooled results see this presentation. For further details on why we interpreted the results as we did, see this document. For rough calculations to calculate the confidence interval, see my spreadsheet.

Meta-analytic result

David went through the meta-analyses; he concluded that all three were reasonably similar, and consistent with an average weight gain from MDA of 0.1kg.



I prefer to think in terms of weight-to-age z-scores, because it seems more likely to me that would be the portable, generalizable unit, which acts as a rough control for age.[3]

In the Campbell review, the **90% confidence interval for weight to age was between 0.00 and 0.10**[4]

Miguel and Kremer result

Miguel and Kremer report height-for-age and weight-for-age z-score differences between Groups 1 and 2. They find insignificant results.[5] Calculating confidence-intervals from the reported mean and standard errors[6] yields a **90% confidence interval between -0.065 and 0.064**. (i.e. smaller but overlapping with Campbell).

---

[3] A change in the weight-to-age z-score reflects a shift in terms of standard deviations in weight for that age. I expect (on priors) that if deworming has an effect, it would be more likely to be a relative rather than absolute effect (e.g. it increases weight by some %, rather than as an absolute number). This means it would likely have less of an absolute effect on younger children.

[4] "In the lower-right of the figure, the bolded numbers and the black diamond present the meta-analytical bottom line: across these 13 trials, mass deworming increased weight by an average 0.05 standard deviations. The aggregate 95% confidence interval stretches from –0.02 to 0.11, just bracketing zero. The final version of the Campbell report expresses the result in physical units: an average gain of 0.09 kg, with a 95% confidence interval stretching from –0.09 kg to +0.28 kg. And so it concludes: "Mass deworming for soil-transmitted helminths with albendazole twice per year compared with controls probably leads to little to no improvement in weight over a period of about 12 months." (How thin the reed)

I've converted 95% to 90% confidence intervals here.

[5] Miguel and Kremer (p173, Table V) report height-for-age Z-score difference as 0.09 (SD 0.09) (significant at 90%) and weight-for-age Z-score difference as 0 (SD 0.04)

[6] See this sheet for calculations

Conversion from nutrition to income

Stephan came up with a best guess for the effect of weight for age on adult earnings from twin studies. I don't feel in a good position to evaluate this without spending a lot of time so I take it as given.[7]

Stephan's central estimate for gain in adult earnings per 1 z-score gain in 0-2 year growth (%) is 3.04%; he notes that this estimate is "*Highly uncertain. I was unable to estimate the effect size of this effectively because there isn't an informative evidence base on it (see stunting report). I assumed an effect size of 2/3 the birth weight effect per z-score. This is based on the general principle (from the animal lit) that the earlier the growth restriction occurs, the more impactful it is*." Stephan early life growth interventions

I interpret this high uncertainty as a lognormal distribution with a 90% CI between 0.1 and 11% with a mean of 3.

---

[7] "I constructed a model of the relationship between early-life growth and adult income based on identical twin studies. According to my model, the income impact of a 0.1 kg increase in weight is an 0.1 percent increase in adult income (this is the weight change reported by meta-analyses of deworming studies with infection intensities similar to Miguel and Kremer).

There is a lot of uncertainty here because 1) The underlying estimate of the growth-income relationship is uncertain and could easily be off by threefold; 2) We don't actually know the mechanism by which deworming may increase adult income, so it's not certain that growth captures it.

These very modest short-term effects on established markers of childhood development, and very small income effect predicted from growth data, make it more difficult to believe that there are large long-term effects on income. Long-term effects would have to occur via mechanisms that are poorly understood, which is certainly possible but seems less likely.

I asked myself this question: Based on the short-term evidence alone (not considering other evidence), what is my 80% credible interval for deworming effects on adult income?

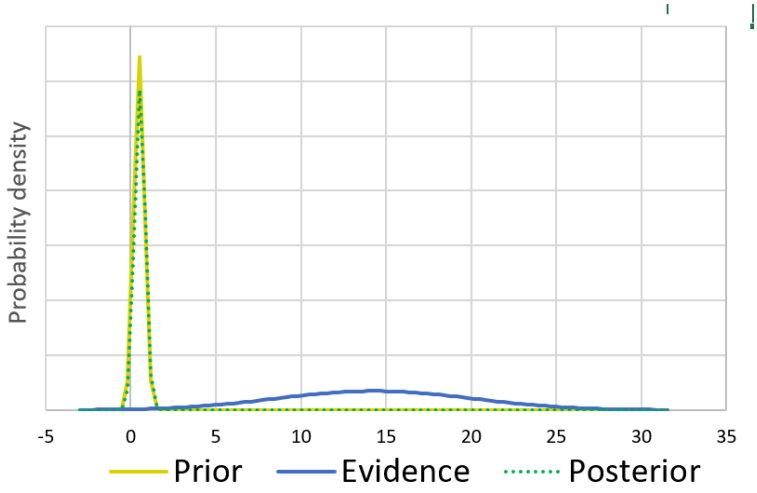
It seems hard to imagine large long-term effects when short-term effects on established measures of child growth/development are small or nonexistent. That said, short-term effects are an indirect correlate of long-term effects and in principle the two don't have to be correlated at all.

Considering this evidence, the mean is 0.1% with my 80% credible interval ranging from -1 to 4%. This implies a skewed distribution with a negative component, which I haven't been able to figure out how to model in David's decider. Instead, I'll model it using a normal distribution with a mean of 0.1% an 80% credible interval between -2.8 and 3%. This is not a great approximation because it puts too much probability mass in the negative, but the mean and SD are similar to the skewed estimate so it should have a similar impact on the posterior mean. Applying the same math as above, the standard deviation is 2.3%."

Putting it all together

Modelling the effect of deworming on income through nutrition (using the Campbell results) gives a **90% credible interval of -0.02% to 0.99%**. I.e. the effect of 2.4 extra years of deworming in Miguel and Kremer was to increase adult earnings between -0.15% and 0.22%.

If we take this as our prior, and update it based on the Miguel and Kremer evidence in David's Bayesian decider, the evidence hardly updates us.



The posterior has a mean of 0.54%, **implying a replicability adjustment of 4%.**

But I think the more sensible conclusion from this is: if weight-for-age plays a role in the causal mechanism between deworming and long run income, it's so small as to be negligible. **The effects of deworming on weight aren't large enough to credibly explain more than a tiny fraction of the long run effects on deworming.**

Cognition

There are two pieces of evidence we might look at to establish how likely it is that improved cognition is a plausible mechanism for the increases in long run earnings in Baird 2016.

The first is the results from various meta-analyses of the effect of deworming on cognition. The second is the results from Baird 2016.

Meta-analytic result

Stephan reviewed the evidence for the impact of deworming on cognitive function. He reviewed Taylor-Robinson, and Welch 2016. Taylor Robinson was a narrative review, while Welch 2016 performed a meta-analysis. Stephan focused on Welch 2016.

When pooling the three studies included in Welch, Stephan concludes that:

"The primary result for general intelligence changes resulting from mass deworming is an SMD of 0.02 with a confidence interval of -0.11 to 0.14. This corresponds to an IQ difference of 0.3 points, with a confidence interval of -1.65 to 2.1 points."

However, on looking deeper into those studies, Stephan concludes that two of the studies (the most positive and the most negative ones) had very low infection prevalence (the most positive) and only a transient treatment effect in helminth infection intensity (the least positive)[8] and so do not provide a valid test of the hypothesis that MDA which is successful at clearing infections improves cognitive outcomes. Looking only at the one study with similar prevalence to Miguel

[8] "Ndibazzaet al. 2012 randomized 2,016 Ugandan children to receive albendazole or placebo quarterly between ages 15 months and 5 years and administered cognitive tests at 5 years(8). Helminth infection rates were low (maximum total infection prevalence was approximately 11 percent in 4-5-year-olds) and albendazole treatment had little impact on infection prevalence16. Unsurprisingly, the impact of the intervention on general intelligence was very null. This study does not appear to be a valid test of the hypothesis, yet it carries 63 percent of the evidence weight of the general intelligence outcome of Welch 2016."
"Nga et al. 2009 randomized 510 Vietnamese children to receive albendazole, and/or multi-micronutrient-fortified biscuits, and/or placebo in a 2X2 factorial design17(9). Ascarisand whipworm infection prevalence and intensity were fairly high at baseline (prevalence was65% for Ascaris, 56% for whipworm, and 6% for hookworm; table 2). Albendazole appears to have been administered only once18, and although it suppressed Ascarisand whipworm levels at two months post-treatment, by four months (when the cognitive tests wereadministered) infection intensity was not very different betweengroups(table 5). Unsurprisingly, the impact of the intervention on general intelligence was null, although there was a trend toward a positive impact. I question the validity of this study as a hypothesis test due to themodest and transient difference in helminth infection intensity between groups. It carries 21 percent of the evidence weight of the general intelligence outcome of Welch 2016."

and Kremer 2004, the mean effect size falls slightly, but the confidence intervals widen.[9] Stephan concludes:

The midpoint estimate of the Watkins general intelligence outcome is 0.01 SDs, which roughly corresponds to a gain of 0.15 IQ points and 0.15 percent consumption. The upper bound of the 95 percent confidence interval for Watkins is 0.31 SDs, which roughly corresponds to a gain of 4.7 IQ points and 4.7 percent consumption. [0.02 [-0.11,0.14].

I'm not sure whether it's reasonable to exclude Nga et al. 2009 (the most positive study), because low effect on worms burden is a real possibility for our top charities. However, as this is the most positive study (even though it didn't successfully clear infections), I'll set that aside.

I have pretty flat priors over these effect sizes (i.e. I don't have any prior reason to think any of these effect sizes is more likely than the other), I interpret these confidence intervals *for IQ* as credible intervals: (i) distributed normally (95% CI, -1.65, 2.1) (ii) distributed normally (95% CI, -4.4,4.7).

I think there are arguments for both probability distributions. I think given the low prevalence of worms in Ndibazza 2012, excluding this study is probably right. But the small effect in Nbidazza et al. 2012 does provide *some small* additional evidence that MDA doesn't have a large effect on cognition in areas with high worm prevalence. Given this, I subjectively place ⅔ weight on (95% CI, -4.4,4.7) and ⅓ weight on (95% CI, -1.65, 2.1).

Combining these, I get a 90% probability distribution between -2.9 and 3.3 points effect on IQ.

IQ to income

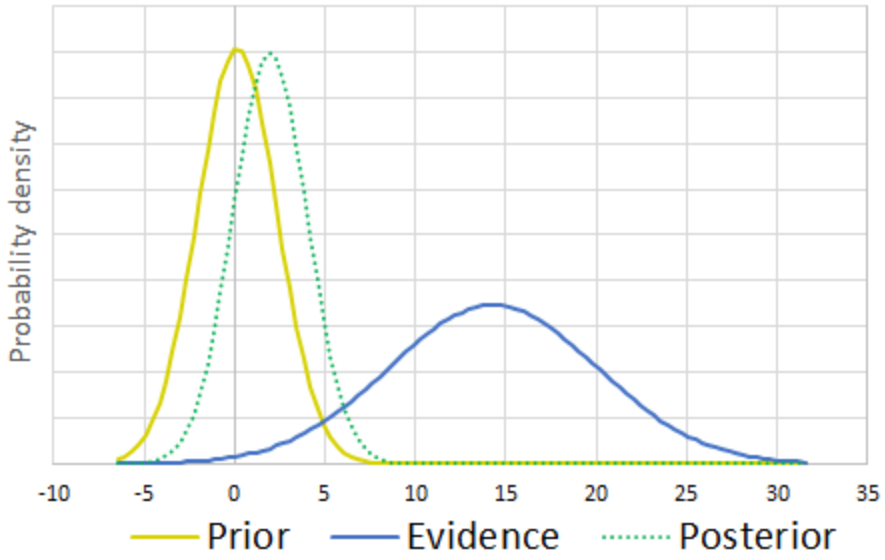
I haven't reviewed this yet, but we use 1 point increase in IQ -> 1% rise in consumption. I think this is worth reviewing more. I've modelled this as a 90% probability distribution between 0 and 2%.

Putting it all together

Combining these two effects gives us a 90% probability distribution of between -3.5% and 3.7% based on the effects of cognition.

[9] Watkins et al. 1996 randomized 246 Guatemalanchildren to receive albendazole or placebo at baseline and 12 weeks, followed by cognitive testing at six months(10). Prevalence and intensity of Ascarisand whipworm infection were high at baseline (91% and 82% prevalence, respectively; no hookworm detected; p. 158 and table 2). Albendazole was very effective at treating Ascaris infection and moderately effective for whipworm(table 2). The outcome was very null, but with fairly wide confidence intervals. This study appears to be an effective test of the hypothesis. It carries 16 percent of the evidence weight of the general intelligence outcome of Welch 2016.

Updating based on the pooled KLPS2 and KLPS3 results provides a modest update. The posterior has a mean of 2%, implying a **replicability adjustment of 14%.**



It still seems fairly unlikely to me (but possible) that improved cognition is the primary mechanism through which deworming affects earnings because:

- The school children in Miguel and Kremer were above the age at which I'd guess their cognitive development would be influenced.
- Baird says that it's unlikely to be this mechanism.

Nevertheless, I take 14% as a lower bound on my replicability adjustment. The prior and evidence distributions have enough overlap that it doesn't seem like a totally crazy pathway.

Years of schooling

Baird 2016 result

Baird 2016 finds a non-significant increase in grades of schooling attained by 2007. 0.150 (0.143).[10] **This translates to a 90% confidence interval between -0.085 and 0.385.**

The Campbell review includes an estimate of school enrolment, but the only study included was Baird. I use the Baird result for increases in grades of schooling, rather than school enrolment, because some of the increase in school enrolment was due to increased repetition of years of schooling.

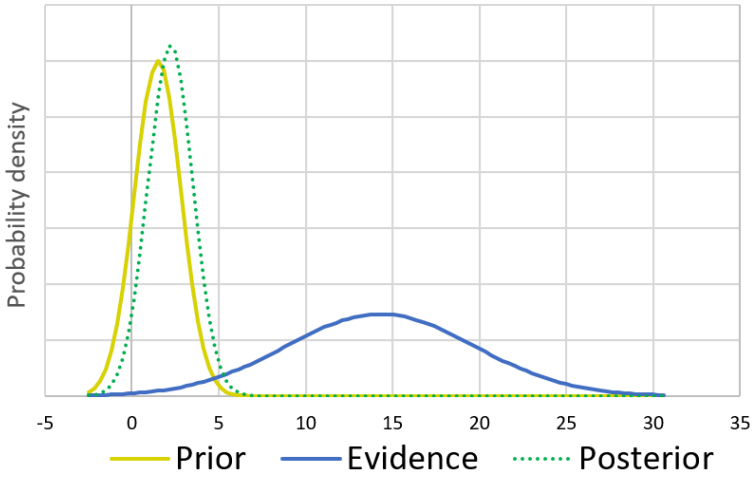
---

10 http://emiguel.econ.berkeley.edu/assets/miguel_research/64/Worms-at-Work_2016-07-12_FINAL_CLEAN.pdf Table 2

I assume an additional year of grade completion increases income by 5 to 15% (90% confidence interval). This is roughly based on my understanding of results from Mincerian regressions, although I haven't reviewed these studies at all.

Putting it all together

The posterior distribution after updating based on the KLPS3/4 evidence, has a mean of 2.22, which implies a replicability adjustment of 16%.



My level of confidence in this type of analysis

I believe the advantages of this type of analysis are:

1. It comes to a concrete replicability adjustment using empirical data and formal Bayesian reasoning.
2. It allows me to understand how particular new data points would update me, by plugging them into my model. For example, the new KLPS4 results will update me by (i) changing the mean expected effect size of deworming on income (ii) giving tighter confidence intervals around the expected effect size, which gives the direct estimates of the effect of deworming on income higher weight in a Bayesian calculation.

I believe the main disadvantages / limitations of this type of analysis are:

1. It only considers a subset of relevant data. In particular, I have not included adjustments for the quality of each study (although I do include adjustments for the precision of the estimates). Instead, my approach is to take the most important data points at face value, and see if there is a consistent theory which allows them to hang together
2. Particularly relevant data points I have not considered (due to time limitations):

    a. The Ozier paper looking at effects of living in a dewormed area on young children who did not receive treatment.
    b. The results from screened trials, suggesting that targeted deworming *does* have an effect on various short term indicators.
    c. The possibility that unprogrammed deworming is common for symptomatic children (and may be more common now than it was in Busia). This would help to explain how MDA could have little to no impact, while screened deworming would.
3. I've modelled cognition, nutrition, and years of schooling as three separate causal pathways that don't interact with each other, which is false. I don't have a good understanding of how they'd interact. Importantly, I don't know how plausible it is that deworming would increase income by improving performance on cognition tests, or increasing years of schooling, without having any major detectable effect on weight gain.
4. I start with flat priors over the effect size of interest. I think this is reasonable: if someone were to ask me, before looking at either the short or long term evidence, what do I believe about the effect of deworming, I think I'd find it challenging to benchmark. Flat priors seem like the best way to model this.
5. I rely on evidence I haven't reviewed in depth to estimate some of the steps in the causal pathway. For example:
    a. Years of schooling to income
    b. Stephan's weight gain to income model
    c. Our estimates of improvements in cognition to income.
6. As a meta-point, I've spent very little time reviewing the deworming evidence compared to David/Alexander/Holden. I think the main value of this work is to synthesise how all the evidence fits together, rather than diving deep into each study.

What would change my mind?

One of the advantages of modelling this explicitly is it tells me what would change my mind, and gives me a model I can plug updates into.

1. **KLPS4 consumption results.** I expect to plug Wave 2 results into my model when they are released to update my cost-effectiveness estimate.
2. **A convincing argument that there's a plausible mechanism that doesn't correlate with short term effects on haemoglobin, weight gain, or years of schooling.** Currently, I'm discounting this possibility because I can't think of any. As I understand it, others at GiveWell think this mechanism might exist.
3. **More precise estimates of effects on years of schooling.** Unfortunately, I think this is only gettable from long term studies.