Summary of 2020 Worm Burden Model

Summary

In 2020, GiveWell introduced a new worm burden adjustment model with updated methodology for measuring worm burden and data on worm prevalence. The main update is that we have shifted from measuring worm burden in terms of average infection intensity to counting the prevalence of "moderate" and "heavy" intensity worm infections (as defined by the World Health Organization (WHO)) and putting more weight on heavy intensity infections than moderate ones. This document summarizes the current model, changes to the model, reasoning for key inputs, and remaining uncertainties.

The 2020 worm burden adjustments represent a substantial update to GiveWell's deworming cost-effectiveness model, with large cost-effectiveness changes for quite a few, but not all, deworming programs. These tables show how cost-effectiveness has changed after incorporating the new worm burden adjustments. These changes in cost-effectiveness are difficult to summarize because there isn't a universal direction to the update.

We are publishing this work now because it represents our current best guesses and is playing a role in our funding recommendations, but it has not yet gone through the full internal review process for GiveWell research. We are also seeking feedback from our top charities that we support for deworming programs.

What is the worm burden adjustment?

The worm burden adjustment is an important component of GiveWell's deworming cost-effectiveness analysis that compares the aggregate baseline worm prevalence for each charity deworming program to worm prevalence in the Miguel & Kremer 2004 (MK) deworming RCT study population, weighted by both worm species and infection intensity. The calculations behind our new worm burden adjustments can be found here. Baseline worm burden was extraordinarily high during MK due, in part, to an El Niño event during the experiment, so all current top charity programs receive a downward adjustment for relative worm burden. These program-level adjustments directly multiply with the present value of lifetime income/consumption benefits from a year of deworming in GiveWell's deworming CEA. The worm burden adjustment, and a 13% replicability adjustment that's common to all programs, are the largest deductions to total benefits in GiveWell's deworming CEAs, and the worm burden adjustment (along with differences in cost per treatment) drives a large share of the difference in cost-effectiveness across different country programs.

Key features of the 2020 worm burden adjustment

A summary narrative of our current modeling assumptions:

- Our model tracks and weights the prevalence of 5 worm species: roundworm (*ascaris*), whipworm (*trichuris*), hookworm, *schistosoma mansoni*, and *schistosoma haematobium*.
    - The first four of these worm species were prevalent during the MK experiment. *S. haematobium* wasn't prevalent during the MK RCT, so we don't have direct evidence of the impact of *S. haematobium* reduction on future earnings. However, we include *S. haematobium* in the model and make downward adjustments to the weight on this species to account for the indirectness of the supporting evidence (discussed further below). The weight to assign to *S. haematobium* in the model is a particularly uncertain assumption.
- We base our worm burden measurements on the thresholds for "moderate" and "heavy" intensity infections defined by the World Health Organization (WHO).
    - Note that this methodology eliminates any weight on "light" intensity worm infections based on a lack of evidence of major morbidity from this category of infections. This is our current best guess, but it's also an important way that we could be wrong, since light infections tend to make up the majority of overall worm prevalence.
- We weight infection intensity by assuming that a heavy intensity worm infection is four times as bad in terms of morbidity as a moderate intensity one.
- We try to consider baseline pre-treatment worm burden rather than current program worm burden, implicitly assuming that worm burden would return to baseline in the counterfactual where deworming ceased. In reality, a true baseline survey of worm prevalence is not available for every program.
- We make analogous calculations for the reference worm burden during the MK experiment, and compare the worm burden of charity programs in the numerator to MK worm burden in the denominator. (When thinking about how different assumptions could change the bottom line, it's important to keep in mind that many changes we could make

to modeling charity worm burdens would also affect how we calculate the reference worm burden.)

- We account for high program outliers among the charity programs we're tracking by setting calculated worm burden adjustments above the 90th percentile equal to the 90th percentile.

Impact on cost-effectiveness

Note that the metrics in this section compare the expected cost-effectiveness of each deworming charity to the expected cost-effectiveness of GiveDirectly's cash transfers, as designated in multiples of cash transfers ("x cash"). More information on benchmarking interventions against cash transfers can be found in this blog post.

The 2020 worm burden adjustments represent a substantial update to GiveWell's deworming cost-effectiveness model, with large cost-effectiveness changes for quite a few, but not all, deworming programs. These changes in cost-effectiveness are difficult to summarize because there isn't a universal direction to the update. They have led to program re-ordering, which is somewhat obscured by looking at changes in overall cost-effectiveness across deworming top charities, since most top charities saw both increases and decreases in cost-effectiveness across programs that the charity supports.

These tables show how cost-effectiveness has changed after incorporating the new worm burden adjustments. Key cost-effectiveness results that may be important for GiveWell's funding decisions in the near future include:

- Our overall END Fund cost-effectiveness estimate is now fairly high at 10.5x cash (previously 4.8x), primarily because our cost-effectiveness estimate for Angola went up sharply with the new worm burden adjustment. That overall estimate may be misleading, though, because we haven't spent time looking into where the END Fund would spend marginal funding, and we haven't updated our cost per treatment analysis for it since 2017.
- The new adjustments didn't have a big impact on Deworm the World's overall cost-effectiveness before updating country weights.
- Sightsavers saw a substantial increase in overall cost-effectiveness from 7.2x cash to 11.2x cash after worm burden updates.
- SCI's overall cost-effectiveness increased slightly to ~10x cash, but there is increased polarization of low/high cost-effectiveness programs after updating country weights. Most of the programs we supported in 2019 increased in cost-effectiveness, but new 2020 funding gaps tended to have low cost-effectiveness.

The worm burden updates have shifted some programs below the level of cost-effectiveness that we currently consider competitive with other opportunities. However, we haven't yet made any program shutdown decisions based on this year's model in cases where we've supported programs in the past, since we want to be cautious about ending programs. We also haven't extended the funding runway beyond 2022 for programs with low cost-effectiveness. There's still a chance that we may positively update on these programs after discussing with charities, particularly because underlying worm burden data may not be precise, and we plan to consider marginal programs on a case-by-case basis in the future.

Key assumptions in the 2020 Worm Burden Model

Key Assumption 1: using 'moderate infection equivalents' to measure worm burden

Overview

This assumption is the biggest update in the 2020 version of the worm burden adjustment compared to the previous version. We used to scale the worm burden adjustment by comparing the *average intensity of worm infection*, which is calculated by summing the number of eggs per gram (epg) counted in all prevalence survey samples and dividing by the number of samples. We're now basing worm burden on the *prevalence of moderate and heavy intensity worm infections* (abbreviated as MHII hereafter) as determined by WHO thresholds categorizing epg counts into infection intensity ranges for each worm species, and putting more weight on the severity of heavy infections by converting them to a unit we call prevalence of 'moderate infection equivalents', according to which a heavy infection is equivalent in severity to four moderate ones.

Note that GiveWell models the benefits of deworming in terms of long-run income increases, and throughout this discussion, we use morbidity attributed to worm infections as a proxy for the cognitive impairment/developmental effects of worm infections that affect future earnings. Because we're not actually sure what mechanisms mediate the link between deworming and measured future income and consumption increases, this assumption could be wrong, but health morbidities seem like the most reasonable available proxy.

Basing our worm burden model on MHII involves the following simplifying assumptions: 1) future economic benefits from deworming were limited to subjects with MHII during the Miguel & Kremer RCT, 2) that this rule also applies in today's deworming programs, and 3) that there are discrete jumps in morbidity/cognitive impairment from increasing levels of worm infection rather than finely tuned differences across the entire distribution of people infected. Additionally, the WHO defines thresholds for MHII at different levels of epg for different worm species, so using the MHII measure is likely to lead to better weighting of the disease burden of different worm species relative to each other.[1] In contrast, the average intensity measure would be more appropriate to use if we believe there is a linear relationship between infection intensity measured by epg in an individual and the benefits of deworming.

[1] WHO handbook, table 7, page 33, https://apps.who.int/iris/handle/10665/42588.

Supporting Evidence

The case for using MHII rather than average intensity hinges on the answers to a few important questions (details follow):

1. Do low intensity worm infections cause morbidity?
   - A: Probably not.
2. Do moderate intensity infections cause morbidity?
   - A: Probably.
3. Do heavy intensity infections lead to worse health consequences than moderate ones?
   - A: Yes. This answer depends somewhat on how infections are distributed within each category, but we're most confident about the potential for major morbidity from heavy worm infections, and it appears that symptoms of heavy infections are more severe.
4. Are the WHO-defined intensity thresholds for moderate and heavy intensity infections roughly right?
   - A: Uncertain for most species; probably yes for hookworm.

The supporting evidence isn't clear-cut, but on balance, we believe that the answers to these questions support shifting to measuring worm burden with MHII. A more detailed discussion of past research on these questions is available in this document (internal to GiveWell). The most important supporting reasons for our conclusions are the following:

- **A key piece of evidence that supports a non-linear relationship between morbidity and intensity.** Smith and Brooker (2010) is a systematic review of the impact of hookworm on anemia. It is the clearest study we found on the relationship between intensity and morbidity. It found that moderate and heavy hookworm infections are significantly associated with reduced hemoglobin, while light hookworm infections were associated with small decreases in hemoglobin that were not statistically significant. A GiveWell analysis evaluating Smith and Brooker's results concluded that the headline effect for light infections was disproportionately small relative to a linear morbidity model, and that heavy intensity infections had 2x the impact on hemoglobin as moderate ones.
- **Types of morbidity caused by worm infections.** It seems likely that the body may compensate for small amounts of the types of symptoms that come from worm infection (blood loss, nutrient malabsorption and loss, and tissue damage).
- **Correlational impact on weight gain.** There was a stronger correlational fit between moderate and heavy intensity STH infection and weight gain than between average intensity and weight gain in a preliminary GiveWell analysis based on the Cochrane review of the relationship between deworming and weight.
- **Evidence on who received economic benefits from the MK RCT.** Correlational evidence from David Roodman's blog post about the generalizability of MK/Baird et al. shows that future income increases are correlated with geographies with high prevalence of MHII.[2]

[2] "Interestingly, restricting to serious cases [measured by MHII] enhances the similarity between the infection curves, just above, and the earlier smoothed graph of earnings impact versus elevation…. [T]he similarity between the prevalence contours and the earnings impact contour shown earlier—high at the

- **Expert opinion.** Experts we talked to and WHO treatment guidelines suggest focusing on MHII. In fact, the WHO's 2030 targets have shifted away from deworming treatment coverage to become entirely focused on intensity of infection based on the salience of MHII for morbidity.[3]
- **Susceptibility of average intensity measurements to outliers.** Average intensity is a coarse measure of infection that can easily be influenced by outliers contributing a large number of epg. It could also give misleading values in contexts where the average prevalence of infection is low, but there are pockets of high prevalence and intensity of infection. Looking at average intensity could be misleading for assessing the proportion of a population that's at risk of major morbidity from worm infections. Prevalence of MHII is more robust to these potential problems.

Uncertainties

Our current best guess is that focusing on MHII better fits the available evidence than using average intensity, but this decision is uncertain. Our major current uncertainties include:

- **Limitations of the evidence.** The supporting evidence on the relationship between worm infection intensity and morbidity is somewhat sparse and weak, and the causal mechanism between reduced morbidity and increased future income is uncertain.
- **How to weight light/moderate/heavy infections relative to each other.** Our decision not to put weight on *any* light infections, including any approaching the moderate threshold, is debatable, as is how much weight to put on heavy infections relative to moderate ones. We're likely to continue to investigate our assumptions, and may at some point decide to assign a small amount of intensity weight to light infections relative to moderate ones (i.e., a weight substantially less than 1).
- **Precision of the intensity definitions.** The WHO-defined intensity thresholds (i.e. "light", "moderate", and "heavy") that we're using to assign intensity weight are broad and rough, and we doubt that there's a major difference in morbidity between individuals who fall just above and just below a particular threshold, though we do expect that there are morbidity differences on average between the categories. Part of the rationale for relying on them is a practical issue of data availability–they're the most commonly available measure of intensity, and we rarely have access to detailed data about how many infections lie close to the thresholds.

---

low elevations and then again around 1260 meters—constitutes circumstantial evidence for a sensible theory: children with the greatest worm burdens benefited most from treatment. Second, that measuring worm load to reflect intensity—moving to the graph just above from the one before—strengthens this resemblance and reinforces the notion of extrapolating from Worms at Work on the basis of intensity (average eggs/gram, not how many kids have any infection)." David Roodman, How Thin the Reed https://blog.givewell.org/2017/01/04/how-thin-the-reed-generalizing-from-worms-at-work/

[3] WHO, Road map for neglected tropical diseases, 2021-2030, p. 11.

Key Assumption 2: weighting infection intensity

The morbidity and mortality attributable to worm infections is uncertain and difficult to quantify.[4] We believe that, on average, heavy infections will cause more severe symptoms than moderate ones, and one of the goals of this investigation was to propose an intensity weight for heavy infections that describes approximately how many times worse the symptoms of a heavy worm infection are than a moderate one. Our current research suggests that a range of **2-5x weight on heavy infections** seems most reasonable, and we've chosen to use a **4x weight.** We calculated this 4x weight on heavy infections based on interpolating Global Burden of Disease (GBD) disability weights assigned to different categories of STH worm infections that are used to calculate DALYs in official GBD estimates and a rough understanding of the process by which GBD calculates DALYs caused by different worm species.

For reference, GiveWell's relative morbidity estimates attribute the following health issues to STH infections at different levels of severity:

- Gastrointestinal symptoms
- Growth deficits
- Anemia (hookworm only)
- More severe gastrointestinal problems from heavy infections

The resulting 4x intensity weight estimate seems intuitively reasonable to us, but remains highly uncertain. We relied largely on the GBD disability weights because GBD data is the only source we're aware of that draws on the evidence base and assigns a numerical estimate to relevant morbidities, but we haven't deeply vetted the GBD estimates. Our calculation also applies several uncertain assumptions to the GBD values, particularly around the relative likelihood of experiencing certain symptoms for individuals with moderate vs. heavy worm infections. In addition, the symptoms that GBD data covers don't include any measures of cognitive impairment from worm infections. This could be a key driver of the developmental effects of worm infections, but the measured health effects may also overlap with and serve as a proxy for cognitive impairments.

Note that in general, higher intensity weights lead to smaller worm burden estimates for most programs. The intuition for this result is that the prevalence of heavy infections was extraordinarily high during MK, and few of today's programs have a high enough baseline prevalence of heavy infections for the effect of upweighting heavy infections in the numerator to exceed upweighting the reference MK worm burden in the denominator.

A major limitation of this work is that the current 4x intensity weight is based on symptoms of STH infection, but we also apply this weighting to schistosomiasis infections. GBD enumerates

---

[4] "The precise amount of morbidity and mortality caused by intestinal nematodes will never be known. This elusiveness is due to the non-specificity of clinical signs, difficulties in parasitological diagnosis and a paucity of reliable and accurate epidemiological data, compounded by the fact that much of the burden is concentrated among countries with weak disease surveillance systems. As a consequence, estimating the global distribution and disease burden continues to be based on informed approximations, using the best available information." Brooker 2010, Section 8.

a broader set of potential morbidities from schistosomiasis, but we expect that it would be substantially more work to investigate these symptoms and decide on an intensity weight specific to schistosomiasis. We haven't prioritized doing this work to date, but it's a likely future priority.

More information on the intensity weight calculation is available in [this document](#) (internal to GiveWell).

Key Assumption 3: species weights

Different worm species cause somewhat different health problems. We include multiple STH and *Schistosoma* species in the worm burden adjustment model, but it's possible that income effects are driven more heavily by some species/symptoms than others. We developed a set of species weights that's used to calculate weighted average worm burden to capture our best guess of the relative importance of each species based on (1) the expected morbidity they cause and (2) the generalizability of the evidence from MK to the role of each species in generating long-run income increases. These species weights are rough and strongly influenced by our intuition and priors.

Below, we summarize the main reasons behind the current weighting scheme for each species (more information available here [internal to GiveWell]):

- **STHs.** We weight the three STH species equally, implicitly using the differing WHO thresholds for moderate and heavy infections to weight the severity of morbidity caused by different species relative to each other.
- ***S. mansoni.*** We assign substantially more weight to *S. mansoni* than to each STH species because there's some evidence that it could be particularly important for morbidity and long-run income increases due to deworming. It has the most direct evidence of long-run impact from the MK trial: MHII of *S. mansoni* were far more prevalent than MHII of other species during the MK experiment, and areas with high *S. mansoni* infection prevalence are the most strongly correlated with long-run income increases.[5] Additionally, our prior is that *S. mansoni* causes more severe morbidity than STH infections, though WHO thresholds likely account for a portion of the difference in morbidity. However, we don't necessarily believe the thresholds fully compensate for the qualitatively different and more varied types of morbidity caused by schistosomiasis infections, especially when mechanisms mediating developmental effects are uncertain.
- ***S. haematobium.*** We have the least confidence in how much weight to assign to *S. haematobium.* Its prevalence was negligible during the MK experiment, so we lack any direct evidence that it causes long-run income increases. Our impression is that experts take morbidity from *S. haematobium* seriously, so it seems plausible that if other worm species lead to developmental benefits, *S. haematobium* could as well. Symptoms associated with *S. haematobium* appear to be at least as severe as, but not identical to,

[5] David Roodman, How Thin the Reed
https://blog.givewell.org/2017/01/04/how-thin-the-reed-generalizing-from-worms-at-work/

those caused by *S. mansoni*.[6] Since *Schistosoma* species are more similar to one another than to STHs, our speculative theoretical modeling assumption is to weight *S. haematobium* and *S. mansoni* equally. However, we arrived at a 15% weight for this species after making a 60% external validity adjustment (i.e. a 40% reduction) to the theoretical weight on *S. haematobium* for indirectness of the supporting evidence, reassigning half of the weight deduction to *S. mansoni* and half to STHs. This assumption is consequential: among charity programs we currently track, *S. haematobium* prevalence contributes 34% of program worm burden adjustments on average and 18% at the median, and it has the potential to substantially lower or raise program adjustments depending on our assumed weight.

This reasoning led to the set of theoretical weights in Table 1, below, and the final adjusted weights after applying the external validity adjustment for *S. haematobium*.

Table 1: Proposed Species Weights for 2020 Worm Burden Model

Species	Initial Theoretical Weight	Final Adjusted Weight (used in model)
S. mansoni	25%	30%
S. haematobium	25%	15%
Ascaris	16.67%	18.33%
Trichuris	16.67%	18.33%
Hookworm	16.67%	18.33%

[6] *S. mansoni* causes intestinal schistosomiasis, while *S. haematobium* causes urogenital schistosomiasis. These conditions are associated with the following symptoms:

"Intestinal schistosomiasis can result in abdominal pain, diarrhoea, and blood in the stool. Liver enlargement is common in advanced cases, and is frequently associated with an accumulation of fluid in the peritoneal cavity and hypertension of the abdominal blood vessels. In such cases there may also be enlargement of the spleen.

The classic sign of urogenital schistosomiasis is haematuria (blood in urine). Fibrosis of the bladder and ureter, and kidney damage are sometimes diagnosed in advanced cases. Bladder cancer is another possible complication in the later stages. In women, urogenital schistosomiasis may present with genital lesions, vaginal bleeding, pain during sexual intercourse, and nodules in the vulva. In men, urogenital schistosomiasis can induce pathology of the seminal vesicles, prostate, and other organs. This disease may also have other long-term irreversible consequences, including infertility." WHO, Schistosomiasis Fact Sheet

Smaller model updates

In addition to the key assumptions above, we made several smaller updates to the new model compared to the previous version:

- **Program data updates.** We updated program worm prevalence data and, in some cases, used new primary sources. While we've tried to improve overall data quality, there are many data limitations, and we still only have moderate confidence in the data for the median program.
- **More detailed MK prevalence data.** We incorporated more detailed data about worm prevalence and intensity during MK provided to us by the authors. This resulted in removing the El Niño adjustment factor from the CEA, since we now have a better (but still imperfect) understanding of how El Niño affected baseline worm burden during the experiment and have incorporated this into the denominator of the worm burden adjustment.
- **Adjusting outliers.** In the previous version of this model, we made a series of rough adjustments to correct for outliers and data uncertainty, including a) mean regression, b) incorporating GBD country prevalence data, and c) winsorizing results above the 90th percentile and below the 10th percentile. We're continuing to winsorize the final model results to move high outliers down to the 90th percentile, but we're no longer putting weight on GBD prevalence data, implementing mean regression, or boosting programs with low prevalence. This change has meaningfully reduced our cost-effectiveness estimates for a few programs with negligible reported prevalence that used to be increased by these adjustments. We plan to consider these programs on a case-by-case basis to understand the accuracy of the worm prevalence data before making funding decisions.

Uncertainties and topics for future investigation

- **How much weight to put on *S. haematobium*.** Making different assumptions about *S. haematobium* has the potential to substantially increase or decrease worm burden adjustments for some programs. We've only done a limited investigation to date, and we'd like to better understand a) morbidity caused by *S. haematobium* and b) the WHO intensity threshold for this species, which could update our inputs. This is likely to be a priority for future investigation.
- **Coinfection.** This worm burden model ignores joint infections with multiple worm species. Each moderate and heavy infection of each species gets counted separately. This is a way our model could be wrong if joint infections are the primary driver of long-run income results, but it isn't a downgrade from our previous model (which also did not account for joint infections).
- **Previous deworming.** Not all charity programs have a true pretreatment baseline prevalence survey, and some programs have experienced many more years of previous deworming than others. In the future, we may adjust baseline worm prevalence to account for previous deworming in order to improve our best guess of counterfactual worm burden.

- **Data imputation.** Infection intensity data is only available for a subset of programs, and we have to impute the prevalence of MHII for remaining programs (~50-60%). Both the new MHII method and the previous average intensity method suffer from lack of intensity data and data imputation issues. We haven't prioritized doing a careful comparison of whether these issues are worse for one method than the other, but our best guess is that the new model isn't a downgrade in data quality.

Supporting documentation

Modeling spreadsheet calculating program worm burden adjustments:
https://docs.google.com/spreadsheets/d/11ADiprPvSNdqHyZdQaqdo2rXRr7gab-f10CX-m5Lct0/edit?usp=sharing
Write-up on average intensity vs. moderate-heavy intensity infection (internal to GiveWell):
https://docs.google.com/document/d/1tbJcv1aKQ9Wrr4cyRteubndgvXDXQhYMoo5Fp90wbiU/edit?usp=sharing
Write-up on weighting heavy vs. moderate infections (internal to GiveWell):
https://docs.google.com/document/d/1Nr_Ipy5J5pciNtQOGO0iqdCR6I-hw4-jO3KkhO56M58/edit
Write-up on calculating reference worm burden during the Miguel & Kremer experiment (internal to GiveWell):
https://docs.google.com/document/d/1paZEYkbSkiwLyBdJOazcpsxKckumPP75vYCbgSUM5xM/edit?usp=sharing
Write-up on species weights (internal to GiveWell):
https://docs.google.com/document/d/1CFRt06bzW_WMc6uCZmP8xuYVU-mEuGf95uZVBaE5c5Y/edit?usp=sharing