

GiveWell's 2020 moral weights

Note to external readers: This is a modified version of a document that was written in October 2020 for an internal audience.

Summary

Moral weights are an important factor in our prioritization decisions (more [here](#)). We created a revised set of moral weights for use in 2020 by combining various inputs, including results from a survey of GiveWell donors that we conducted this year. See [here](#) for the new moral weights. This year's work builds on our previous work; in particular, in 2017 we analyzed [here](#) how our approach (at the time) compared to others', and in 2019 we wrote [here](#) about research that we supported.

What do we need from our moral weights?

We need a way to value the main outcomes of the programs we're considering funding, which include deaths averted at various ages, increases in consumption, and more (e.g. stillbirths averted) as we expand the scope of programs we consider.

What methods have we tried or considered?

We've considered disability-adjusted life-years (DALYs); we used to use moral weights assigned collectively by GiveWell staff; and last year we incorporated results from an IDinsight study we funded on the preferences of people who are demographically similar to the people served by our recommended programs. Each of these methods has limitation(s) preventing us from wanting to rely on it entirely. More [here](#).

What did we do this year?

We asked donors for their preferences between averting different numbers of deaths at different ages. We then aggregated those preferences into a coherent set of moral weights across ages, and combined that set of weights with other inputs (e.g. proxies for GiveWell staff views and for IDinsight survey results). We're now using this combined set of weights for deaths averted at different ages, and maintaining approximately the same relative weights for deaths averted vs. increases in consumption as we used last year. More [here](#).

We view this year's weights not as a finished product, but as building on work we've done in the past, to try to come to a more robust "answer" to this unanswerable question. We use the donor survey results as a key input in this year's set of moral weights, but this isn't much of a departure from weights we've previously used – they're similar to previous aggregated GiveWell staff weights and to measurements of years of life lost, and they're similar to the weights we used last year which were informed by the IDinsight study.

What will we do in the future?

Our work on moral weights isn't finished, and in particular we anticipate doing more work to gather information from others (e.g., populations impacted by our programs and our donors) so that we can continue to refine our moral weights. More detail [here](#).

Why incorporate donor weights?

Donor weights tell us what the users of our research think about moral weights and give us granular information that enables us to create a full set of values for averting deaths at different ages. They were also relatively quick and easy to gather. We incorporate them alongside other inputs, like those from the IDinsight study. More [here](#).

How does this change our bottom line?

The 2020 moral weights make AMF, HKI, and Malaria Consortium each look 4% to 13% more cost-effective overall, relative to GiveDirectly. They leave the cost-effectiveness of our deworming charities relative to GiveDirectly nearly identical. Functionally, this is because our new weights value young children slightly more highly relative to people over the age of five, and essentially leave the relationship between averting a death and increasing consumption unchanged. More [here](#).

[GiveWell's 2020 moral weights](#)

[Summary](#)

[Why do we need moral weights?](#)

[What do we want our moral weights to achieve?](#)

[What do we want to satisfy?](#)

[What approaches have we tried or looked into, and what have the challenges been?](#)

[Disability-adjusted life-years \(DALYs\)](#)

[Moral weights assigned collectively by GiveWell staff](#)

[Moral weights based on survey of low-income people in Ghana and Kenya](#)

[What are we working on now?](#)

[Deaths averted at different ages](#)

[Consumption](#)

[Years lived with disease/disability \(YLDs\)](#)

[What are the results?](#)

[Headline results compared to 2019](#)

[Graph of age-weighted results](#)

[Comparison with different age-weighting approaches](#)

[Why incorporate donor preferences?](#)

[What will we do in the future?](#)

Why do we need moral weights?

If you could either prevent the death of an adult or prevent the death of a newborn, which would you choose? What if you could either (a) give 25 very poor households a cash transfer that would double their income for a year or (b) prevent one death of a child from malaria?

We need to compare each program we consider funding to the other programs we may fund instead, which means we have to set moral weights (i.e. subjective valuations) for the outcomes of each program, e.g. saving the life of a toddler, doubling the consumption of a household of four for one year, or averting a stillbirth. These values, even if they are uncertain, enable us to understand whether programs are likely to help people a similar amount per dollar donated or whether one is likely to be much more helpful to the people our recommended programs serve.

Anyone deciding to donate to one charity over another is implicitly using moral weights, even if they aren't explicitly engaging with them. We include moral weights in our cost-effectiveness analyses because they are an important part of any giving decision and we think it is valuable to be transparent about them. Engaging with questions like this is difficult and uncomfortable, but necessary.

What do we want our moral weights to achieve?

What do we want to satisfy?

We're trying to help people as much as possible. Part of what that entails is figuring out how much different outcomes help people, relative to each other.

We don't think there's an objective right answer that we could come to here; we aren't going to "solve" moral weights. We want our moral weights to be good enough at approximating what's most helpful to people that they steer us toward recommending the best giving opportunities we can.

We want to do as much empirically as we can and reserve moral weights for decisions where we think we're not able to make sound empirical judgments. For example, in evaluating an educational program, we'd want to do the empirical work to understand how a year of education affects income/consumption.

We also want our moral weights to be broadly in keeping with what the users of our research – i.e. donors – would expect. This doesn't mean that our moral weights need to be the same as donors', but that if they're different, we need to communicate clearly about that.

Finally, we don't want our moral weights to swing wildly from year to year, making it difficult for charities to engage with us and causing wild swings in how promising we think a program is.

What approaches have we tried or looked into, and what have the challenges been?

In brief (more relevant considerations are listed below):

- Disability-adjusted life-years (DALYs) are widely used, but they don't consider non-health benefits, such as increases in income/consumption.
- GiveWell staff-assigned moral weights allow us to hear from people who are very familiar with the question at hand, but can vary year to year as staff composition changes and also rely on the views of a small number of people.
- A survey of low-income people in Ghana and Kenya conducted by IDinsight gives us insight into the preferences of people similar to those who are affected by our recommended programs, in theory, but we don't believe the results represent a true, credible set of preferences. The results aren't granular enough for our purposes and are also highly uncertain.

Disability-adjusted life-years (DALYs)

[DALYs](#) are a widely used metric in which one DALY represents one lost year of healthy life.

Benefits of using DALYs:

- DALYs are a commonly used metric in the public health community – that makes them easy to explain.
- Because DALYs are widely used, disability weights have already been assigned for many diseases and health conditions.

Reason not to rely on DALYs:

- DALYs don't evaluate income/consumption benefits or any benefits that aren't strictly health-related, such as avoiding grief from deaths or improving people's general sense of wellbeing via making contraception accessible; DALYs aren't intended to account for anything beyond morbidity and mortality. They also don't provide a weight for stillbirths. We can't rely solely on DALYs for this reason, though DALYs are still sometimes useful as an input, especially in early-stage cost-effectiveness analysis.
- DALYs assume a disability weight for a given condition that might be inconsistent with our view (or the views of people impacted by our programs or our donors).¹

¹ For example, though it has changed, in 2008, the disability weight for blindness was .6, i.e., [being blind was 60% as bad as dying](#).

- DALYs assume that preventing the deaths of people with longer remaining life expectancies is always more valuable, when in practice many people indicate a preference for preventing the death of an older child over the death of a neonate.²

Moral weights assigned collectively by GiveWell staff

For several years, we invited all staff to come up with a set of moral weights. Then, we took the median of the results of using each staff member's moral weights and used that as our final cost-effectiveness estimate.

Benefit of staff-assigned moral weights:

- Staff are very familiar with the question we're trying to answer, so we're getting answers that come from a place of respondents being engaged with what we're doing and how we'll use these figures.

Reasons not to rely on staff-assigned moral weights:

- We have fairly few staff, compared to the number of people who can be surveyed via other methods.
- Staff don't have a unique ability to make these moral judgments. Staff also have limited insight into what the lives of people impacted by our recommendations are actually like, and what would be the most helpful to them.
- Staff-assigned moral weights are hard for charities to predict, in that there can be wide swings based on changes in staff composition.
- In past years, staff engaged to varying degrees, and then all those responses were aggregated without weighting for level of engagement.

Moral weights based on survey of low-income people in Ghana and Kenya

In 2019, we recommended funding for an IDinsight to survey approximately 2000 people in Ghana and Kenya living in poverty about their preferences around averting deaths at different ages and around trading off averting a death with increasing consumption. We supported this survey because we thought it would be helpful to learn about the moral weights of people who are from similar backgrounds to the people served by programs we fund, i.e. people living in poverty, most of whom are in Africa. You can read more about this project [here](#).

Benefit of moral weights from the IDinsight survey:

- The people we're trying to help with our work have the most insight into what would actually be most helpful to them. To the extent that what's most helpful is context-dependent, this is really important. By learning what respondents value and taking that into account, we can avoid the pitfall of being an outside funder pushing its own views without having local context.

² This applies to GiveWell staff and donors, as well as to the results of the Mechanical Turk survey referenced [here](#). The IDinsight survey did not ask about neonates.

- It represents opinions on these topics from approximately 2000 people, which is much more raw data than we'd previously had on this topic.

Because we talked about the IDinsight survey in detail last year and in some ways it provided us with more data than we'd ever had before, it's also worth describing in detail the factors that prevent us from relying on it for our final moral weights. Limitations are also discussed [here](#).

Reasons not to rely on moral weights from the IDinsight survey:

- Overall, we don't really trust the results as providing a coherent and accurate picture of people's preferences. The questions were quantitatively complicated and very hard to answer, the sample size was limited for some of the questions, and social desirability bias may have played a major role in people's answers.
- It doesn't cover all of the outcomes on which we need moral weights, disaggregated in the way that's most useful to us.
 - The headline results are only deaths of over-5-year-olds vs. 5-year-olds-and-under (because that's what we use in our HKI, AMF, and Malaria Consortium cost-effectiveness analyses). We'd like to be able to consider the ages at which deaths occur at a more granular level, and also have specific moral weights for outcomes like stillbirths and deaths of neonates. The biggest uncertainty we have is around the relative value of preventing deaths at very young ages.
 - There is data for more granular age levels, but the sample sizes are small and the data doesn't form a coherent picture (e.g., it shows 7- and 8-year-olds valued much more highly than any other ages, including 6- and 9-year-olds, and we aren't aware of an explanation for why that would be a true preference as opposed to a fluke of small sample sizes).³
- Lots of respondents placed extremely high value (over \$10 million) on life. That doesn't necessarily imply lack of engagement with the question – it might be a true preference – but it's not helpful for resource allocation. It's an extraordinary result without extraordinary evidence to support it. If we believed this was an "accurate" representation of how the world should work, we expect that it would imply that low- and middle-income country governments and other institutions should be acting very differently than they currently are.⁴ Possible explanations for this result include social desirability bias and viewing life as too sacred to be valued at any amount of money.
- A relatively minor point: We don't fully understand the underlying rationale behind the valuations, so we don't know whether it might rely on opinions with which we fundamentally disagree.

³ See [pg. 69](#).

⁴ See [here](#) for how different high-income country governments currently value a statistical life.

What are we working on now?

This year, we use the IDinsight survey, DALYs, and GiveWell staff's past moral weights, alongside a new piece of information – donor preferences for trading off lives saved at different ages – to generate our 2020 set of moral weights.

We surveyed about 70 of our largest donors with questions about who they would prefer to save at different ages. We also asked about their reasoning. Then, we converted those responses to a set of moral weights.

Deaths averted at different ages

We combined the donor preferences with years of life lost (YLLs) and with past staff weights, using years of life lost as a proxy for the IDinsight survey (because they look broadly similar, but the IDinsight survey results are too jagged to use wholesale) and using James Snowden's 2018 moral weights as a proxy for GiveWell staff (because his weights were near the median and he had provided a complete set of age groups).

We weighted these sources as follows:

- 60% weight on donor responses
- 10% weight on James Snowden's 2018 weights (as a proxy for 2018 GiveWell staff)
- 30% weight on YLLs (both as a commonly-used metric itself and as a proxy for the IDinsight survey)

Donor responses received the majority of the weight primarily because in our view, that curve represents the most plausible set of weights. In particular, the other two do not distinguish between early neonatal deaths and deaths at age one, which is an age range we're especially interested in.

We use donor weights as our result for stillbirths because we don't have data from the other sources on stillbirths.

Consumption

We didn't get any additional information on the value of consumption this year, and so decided to keep the value of a year of doubled consumption the same (i.e. 1/100th) relative to the value of averting the deaths of under-5s and 5s-and-over from malaria. As the values of an under-5 and 5-and-over death averted from malaria are no longer the same, we take the average.

Years lived with disease/disability (YLDs)

To get a benchmark estimate for YLDs, we divided [the average value of averting a death over the age of 10] by [the average years of life remaining for people over the age of 10]. We chose to only include people over the age of 10 because our 2020 moral weights appear to "downweight" preventing deaths of young children relative to their years of life lost.

What are the results?

Headline results compared to 2019

Here's how the 2020 results compare to last year's weights:

	2019	2020
Value of doubling consumption for one person for one year	1	1
Value of preventing one under-5 death from malaria	100	116.9
Value of preventing one 5-and-over death from malaria	100	83.1
Value of preventing one under-5 death from vitamin A deficiency	100	118.4
Value of averting one stillbirth (1 month before birth)	N/A	33.4
Value of averting one neonatal death from syphilis	70	84.0
Value of averting one year of life lived with disease/disability (YLD)	3.3	2.3

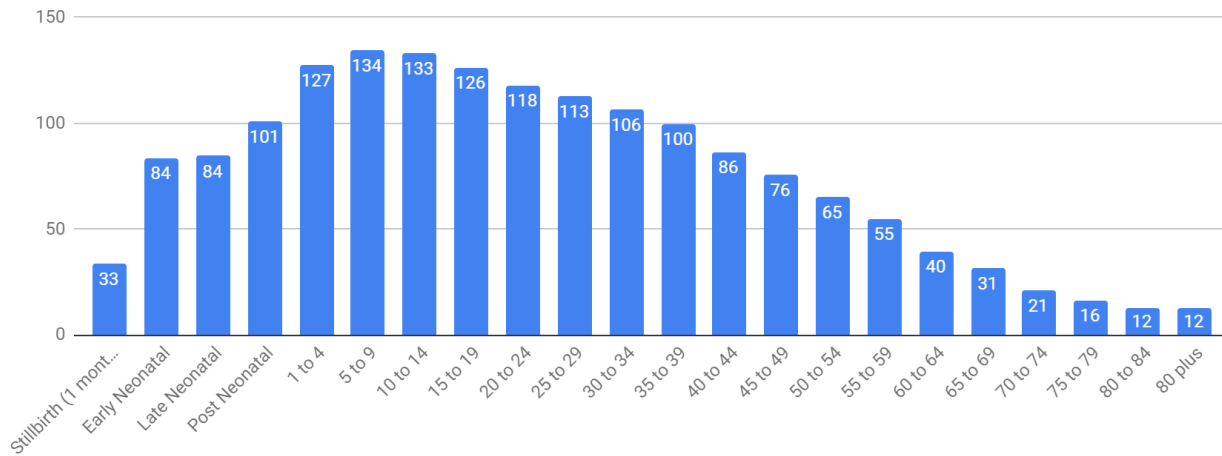
November 2021 update: In 2021, we completed a small update to the 2020 estimates used in our CEAs for our top charities, using data from the 2019 version of the Institute for Health Metrics and Evaluation's Global Burden of Disease project, rather than the 2017 version.

	2019	2020
Value of doubling consumption for one person for one year	1	1
Value of preventing one under-5 death from malaria	100	116.3
Value of preventing one 5-and-over death from malaria	100	73.2
Value of preventing one under-5 death from vitamin A deficiency	100	118.7

Graph of age-weighted results

Here are the results of our aggregation of donor responses, YLLs, and James' 2018 moral weights, where one unit = one year of doubled consumption for one person:

Deaths at different ages in units of doubling consumption



The relative weights here come from the methodology described above, and the absolute weights were determined by keeping it so that the average of [the value of averting the death of an under-5 from malaria] and [the value of averting the death of a 5-and-over from malaria] is 100. (Last year we had set both those values to 100, and we are maintaining the same relative value of consumption and lives saved.)

These results look sensible to us. We're least certain about the value of averting deaths at very young ages and stillbirths. If those values became decision-relevant for a grant, such as a neonatal health program, we might revisit these weights or consider setting aside a pot of funding for grants that satisfy "other reasonable moral weights."

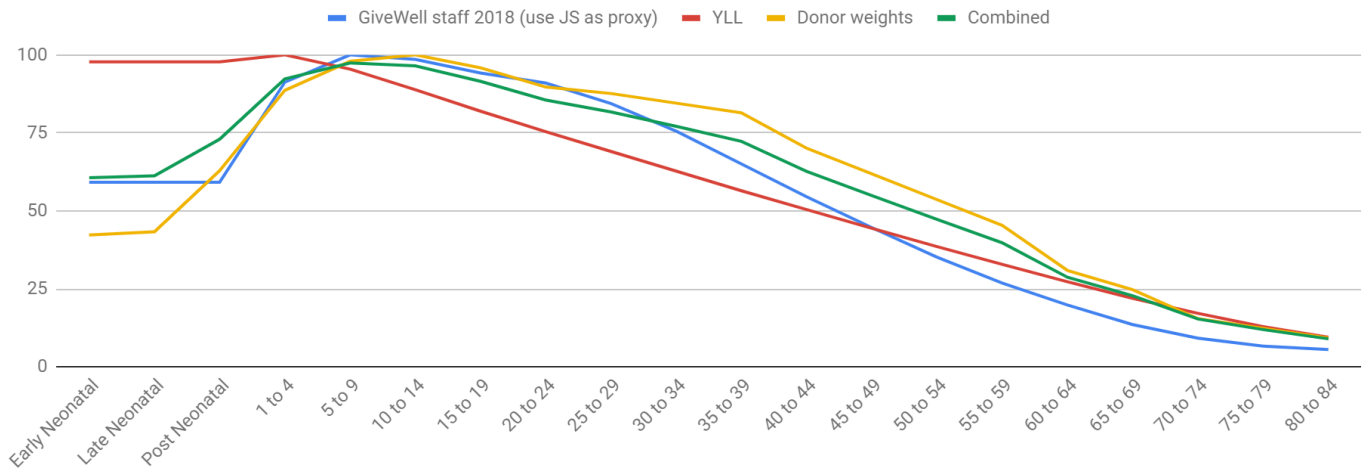
The 2020 moral weights make AMF, Malaria Consortium, and HKI each look 4% to 13% more cost-effective overall, relative to GiveDirectly. They leave the cost-effectiveness of our deworming charities relative to GiveDirectly nearly identical.

Comparison with different age-weighting approaches

The below line graph shows how the different sets of moral weights compare to one another, with the green line representing what we're now using.⁵

⁵ Note that this graph is indexed to 100 and the above chart isn't, so the absolute figures don't match. Stillbirths are excluded because only donor weights provide a value for stillbirths.

Relative moral weights at different ages under different perspectives (indexed to 100)



These curves take similar shapes. James' 2018 weights and donor weights are very similar, with the most substantial difference being that James didn't distinguish between different ages within the first year of life, whereas that's explicitly something we asked donors about. There are two main differences between donor weights and YLLs:

1. YLLs lead to a sharp discontinuity between stillbirths, even late in pregnancy, which receive zero weight, and early neonatal deaths, which receive a lot of weight. Donor weights create a gradual increase from fetal development through early childhood – that seems intuitively more plausible to us than such a sharp discontinuity at birth. (Neonates are still valued substantially more highly than averted stillbirths in our 2020 weights.)
2. There's a small "bump" from around ages 10 to 60 (most pronounced at ages 20 to 40), where averting deaths is valued more highly than you would expect based purely on YLLs. This intuitively makes sense as these are approximately the years of age in which people are most likely to have dependents.

We don't have enough information to know how IDinsight survey respondents would respond to these two divergences from the YLL framework.

Last year, when we received results from the IDinsight study, we updated from the results of that study in two important ways:

1. Valuing averting a death more highly compared to increasing income.⁶

⁶ Last year's weights, like this year's weights, value saving the life of an under-5-year-old 100 times as highly as doubling an individual's consumption for one year. Previously, the median staff member had valued saving the life of an under-5-year-old only 50 times as highly as doubling consumption. (The way this factored into the cost-effectiveness was more complicated than just taking the median staff member's input.)

2. Valuing averting the deaths of young children more highly compared to averting the deaths of older children and adults.⁷

These changes are maintained in our 2020 moral weights. The first is kept identical to last year's value, and on the second, our 2020 approach has led us to update further in favor of valuing averting the deaths of young children over older people (at least for malaria-related deaths – different age distributions might yield different results).

Why incorporate donor preferences?

Each time we update moral weights, we try to make our weights more reflective of what we think actually helps people the most, while updating in a way that avoids major swings from year to year and brings us closer to how we think our weights will look in the future.

Like the approaches discussed above, the donor weights from this survey have pros and cons.

Benefits of incorporating donor survey results:

- Unlike other approaches, the donor survey gives us a value for averting a stillbirth and values for the relevant age ranges at a more granular level. We were able to fairly quickly and easily get this information, compared to how difficult and time-consuming it would have been to get this information from another survey similar to the IDInsight survey.
- We want donors to know what they're getting when they give to GiveWell. Surveying donors served the double purpose of giving us helpful new data and confirming whether our moral weights would be intuitive to donors.
- ~70 donors is a larger sample size than just asking GiveWell's staff would have yielded.

Limitations of relying on donor survey results:

- Donors have limited information about the lives of people who are impacted by our recommended programs, so to the extent that context is important, donors may not have it. (This may apply more to income vs. health trade-offs, which we didn't ask donors to make in this round of surveying. Age weights seem less likely to rely heavily on context.)
- We don't believe the group of donors we surveyed is very diverse (across characteristics like race, gender, income, and country of origin) which could influence results. The vast majority of the donors we surveyed are men,⁸ and people of different genders could

⁷ Last year's weights valued averting the deaths of both children under 5 and people 5-and-over the same. Previously, the median staff member had valued averting the death of a child under 5 only half as highly as averting the death of a person age 5 or older. (The way this factored into the cost-effectiveness was more complicated than just taking the median staff member's input.)

⁸ Anecdotally but confidently, a sizable majority of our major donors are men, so this seems plausibly representative of our major donors.

especially have different intuitions about the value of averting stillbirths and the deaths of neonates.⁹

What will we do in the future?

Our current set of moral weights represents our current "best guess," but we plan to refine it, while avoiding large swings from year to year.

There are several perspectives from which we can seek moral weights: program participants in various communities, donors, other experts like philosophers and development practitioners. The primary challenge we face is in designing a survey that gets us the information we need. As we have the capacity to do more of this, we want to continue to take this on.

Here are some of the things we would like to do:

- Do more donor surveying for other moral weights, like increases in income/consumption vs. deaths averted. We may also test whether giving donors more information affects their responses.
- Ask IDinsight for some of the raw data from the survey they conducted – we currently have the results – to see what we can make of it and whether it's informative.
- Do more surveys of populations similar to those served by the programs we fund, like the IDinsight survey, to see if we can get more usable information if we structure the survey differently, such as by asking about young children in a more granular way.

⁹ A quick analysis of the responses of men vs. women didn't indicate that we should upweight stillbirths and the deaths of neonates to account for different preferences across genders, but there were so few women in the sample that we can't say with confidence that the results don't depend on gender.