The elephant in the bednet: the importance of philosophy when choosing between extending and improving lives

Michael Plant, Joel McGuire, and Samuel Dupret

November 2022

Happier Lives Institute

Contents

Summary[1]

How should we compare the value of extending lives to improving lives? Doing so requires us to make various philosophical assumptions, either implicitly or explicitly. But these choices are rarely acknowledged or discussed by decision-makers, all of them are controversial, and they have significant implications for how resources should be distributed.

We set out two crucial philosophical issues: (A) an account of the badness of death, how to determine the relative value of deaths at different ages, and (B) locating the neutral point, the place on the wellbeing scale at which life is neither good nor bad for someone. We then illustrate how different choices for (A) and (B) alter the cost-effectiveness of three charities which operate in low-income countries, provide different interventions, and are considered to be some of the most cost-effective ways to help others: Against Malaria Foundation (insecticide-treated nets), GiveDirectly (cash transfers), and StrongMinds (group therapy for depression). We assess all three in terms of wellbeing-adjusted life years (WELLBYs) and explain why we do not, and cannot, use standard health metrics (QALYs and DALYs) for this purpose. We show how much cost-effectiveness changes by shifting from one extreme of (reasonable) opinion to the other. At one end, AMF is 1.3x better than StrongMinds. At the other, StrongMinds is 12x better than AMF. We do not advocate for any particular view. Our aim is simply to show that these philosophical choices are decision-relevant and merit further discussion.

Our results are displayed in the chart below, which plots the cost-effectiveness of the three charities in WELLBYs/$1,000.[2]
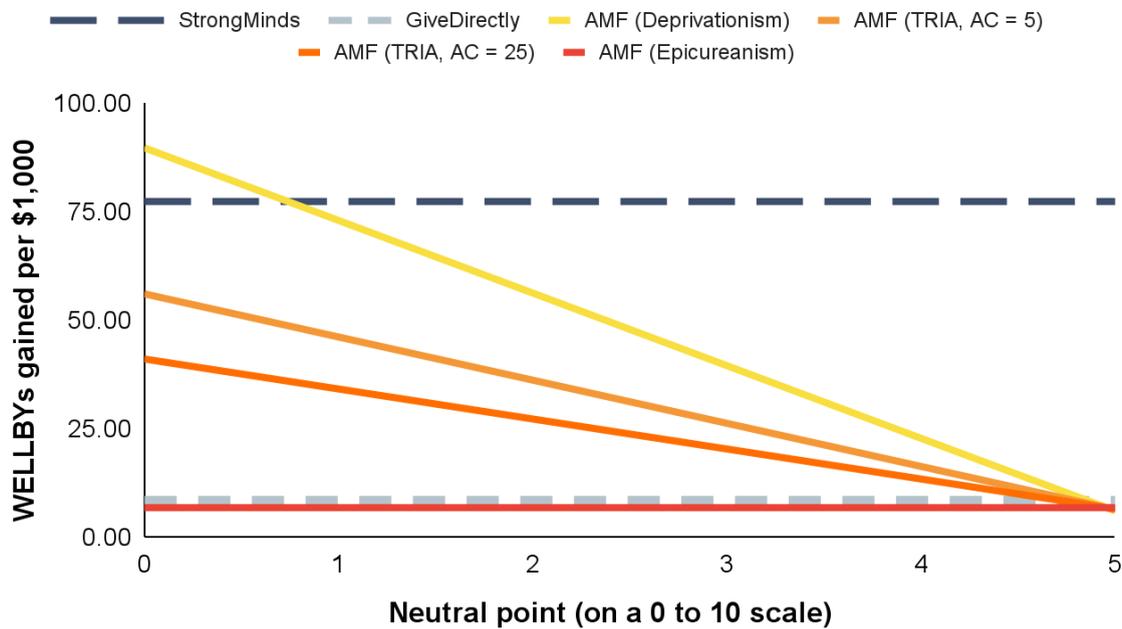
StrongMinds and GiveDirectly are represented with flat, dashed lines because their cost-effectiveness does not change under the different assumptions. The changes in AMF's cost-effectiveness are a result of two varying factors. One is using different *accounts of the badness of death*, that is, ways to assign value to saving lives at different ages; these three accounts go by unintuitive names in the philosophical literature, so we've put a slogan in brackets after each one to clarify their differences: *deprivationism* (prioritise the youngest), the *time-relative interest account* (prioritise older children over infants), and *Epicureanism* (death isn't bad for anyone – prioritise living well, not living long). We also consider including two variants of the *time-relative interest account* (*TRIA*); on these, life has a maximum value at the ages of either 5 or 25. The other factor is where to locate the *neutral point*, the place at which someone has overall zero wellbeing, on a 0-10 life satisfaction scale; we assess that as being at each location between 0/10 and 5/10. As you can see, AMF's cost-effectiveness changes a lot. It is only more cost-effective than StrongMinds if you adopt *deprivationism* and place the neutral point below 1.

---

[2] 1 WELLBY is equivalent to a 1-point increase on a 0-10 life satisfaction scale for one year. For example, going from 5/10 to 6/10 for one year.

1. Introduction

How should we compare the value of extending lives to improving lives? Let's focus our minds with a real choice. On current estimates, for around $4,500, you can expect to save one child's life by providing insecticide-treated nets (ITNs). Alternatively, that sum could provide a $1,000 cash transfer to four-and-a-half families living in extreme poverty ($1,000 is about a year's household income).[3] The cost of both choices is the same, but the outcomes differ. Which one will do the most good?

This is a difficult and discomforting ethical question. How might we answer it? And how much would different answers change the priorities?

There are various methods we might draw on to make the comparison, such as health measures (DALYs and QALYs), subjective wellbeing measures (WELLBYs), or combining the judgements of experts, beneficiaries, and donors (e.g. GiveWell's moral weights approach). However, whichever method we use, we must make some philosophical assumptions.

There are two philosophical issues that make a significant difference to the relative value of improving lives to extending lives. Specifically, these are: (A) an account of the badness of death, and (B) the location of the neutral point on a wellbeing scale. How we settle these two questions can change whether it's better to extend or improve lives.

Although there is philosophical literature on these topics,[4] they have received surprisingly little attention within the effective altruism community,[5] particularly given their potentially significant influence on decision-making. To address this issue, the charity evaluator GiveWell assigns 'moral weights' to deaths at different ages compared to doublings of consumption (see Figure 1). For

---

[3] This is the case for household income of the poorest 10% in Guinea, Chad, Uganda, Togo, Nigeria, and the Democratic Republic of Congo (see row 258 of "Modified" tab for calculations).

[4] See Gamlud and Solberg (2019) for a recent, excellent compendium on the philosophical issues applied to health policy that we draw on several times in this report, Luper (2019) for a general overview and this PhilPapers archive for a list of hundreds of paper on the topic.

[5] For mentions in an effective altruist context, see Plant (2016; 2019, Chapter 3), who sketches arguments that philosophical issues could make a very substantial difference to choosing between life-extending and life-improving interventions. Cotra (2016), writing for GiveWell, responds and objects to Plant (2016). Donaldson et al. (2020) estimate the relative values of doubling consumption and saving the life of a 5-year-old using a wellbeing-adjusted life year (WELLBY) approach. They emphasise that switching from one view of the badness of death (*deprivationism*) to another (the *time-relative interest account*) would reduce the value of saving 5-year-olds by a factor of five, a substantial difference. Donaldson et al. (2020) does not estimate the cost-effectiveness of any charities or say much about the different philosophical theories and how the cost-effectiveness estimates of charities change under these assumptions. These are things we do in this report.

instance, saving the life of a 5-to-9-year-old is deemed to have the same value as 134 years of doubled consumption. GiveWell derives these numbers by surveying its staff, its donors, and its beneficiaries and then weighting their answers (GiveWell, 2020). However, there is no public discussion of why - for what reasons, on what grounds - someone might choose one moral weight rather than another. Are these weights sensible? How could we tell? Nor is there discussion of how much difference this choice makes. Do ethical choices matter, or do you get essentially the same answer whatever 'reasonable' theoretical position you take?[6]

Figure 1: GiveWell's disvalue of death at different ages in units of doubling consumption
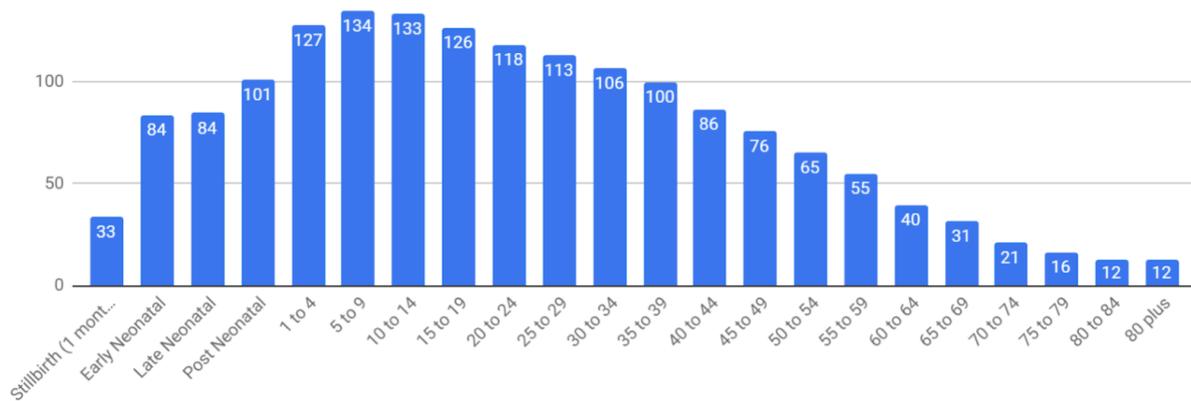
Note: This graph comes from GiveWell (2020). Plant (2022a) discusses these weights in more depth.

This report is written for those who must make difficult choices between extending and improving lives and would like to have a better understanding of: (1) the theoretical choices they must make, (2) some of the reasons for making one choice rather than another; and (3) how much of a difference this could make. Our hope is that decision-makers will gain a better understanding of the issues and priorities that influence their decision rather than relying on opaque methods and recommendations made by others.

In section 2, we provide the theoretical background. We outline the two key issues, the options within them, and the advantages and disadvantages of those options.

In section 3, we put this theoretical machinery to work and show how the philosophical factors are decision-relevant by providing a novel cost-effectiveness comparison of three top charities.

In section 4, we set out our key uncertainties and directions for future research.

In section 5, we briefly conclude.

---

[6] Note some discussion of GiveWell's moral weights on this recent EA forum post.

2. Philosophical issues in the valuation of extending lives

The standard metrics for combining quality and quantity of life are quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs), two measures of health that were developed in the latter third of the 20th century (Sassi, 2006). On the QALY model, health states are modelled as values between 0 (death) and 1 (full health). For instance, two years lived at 0.5 would have a value of 1 QALY, as would one year of perfect health, and so on. This allows different health improvements to be quantified and compared in the same units. The DALY is essentially the same, but it is a measure of loss, rather than gain: 1 represents full disability and 0 represents no disability.[7] A principal motivation for the DALY was the Global Burden of Disease (GBD) study, which sought to quantify health trends and challenges worldwide (Salomon et al., 2012).

In the earlier (1990-2010) versions of the GBD, deaths were considered to be the worst for 10-year-olds. In the 2010-2015 version, the deaths of infants were considered worse than those of 10-year-olds, but stillbirths did not generate any disease burden (see Gamlud and Solberg, 2019, pp. 6-12 and references therein). This raises the question, 'what is the worst age to die?', and highlights issues that are referred to in the philosophical literature as *the badness of death* (Nagel, 1970; Feldman, 1992; Broome, 2004; Gamlud and Solberg, 2019). In section 2.1, we set out the three main accounts - *deprivationism*, the *time-relative interest account*, and *Epicureanism* - and show how the relative values of saving a life at different ages vary substantially between, and even within, them. In section 2.2, we turn to a further issue, that of the *neutral point*.

2.1 Accounts of the badness of death

Deprivationism (prioritise the youngest)

The more recent version of the GBD implies a *deprivationist* account of the badness of death.[8] On this, the badness of death consists in the difference between the wellbeing you *actually* had compared to how much you *would* have had if you'd lived longer. That's why it's better to save the youngest lives: they have more to lose than older ones. Two leading proponents are Broome (2004) and Bradley (2009).

---

[7] For our purposes, the differences are not important. See Sassi, 2006 for a discussion.

[8] This is sometimes called the *life comparative account*. Arguably, this would be a more appropriate name, but we stick with the conventions of the literature. The rival view to (what we've called) *deprivationism* is the *time-relative interest account*, but both agree that death is bad because it *deprives* you of the goods of life. Hence, to call the former the *life comparative account* would make it easier to distinguish it from the latter.

Formulaically, the badness of death on *deprivationism* is:

$$wellbeing\ lost = wellbeing\ level \times years\ of\ life\ lost$$

For example, if someone is expected to live 50 years at wellbeing level 1, then the badness of their death would be equivalent to 50 'units' of wellbeing lost. We return to the issue of determining 'wellbeing levels', particularly in relation to a 'zero' level, in section 2.2.

The counter-intuitive aspect of this view is that it requires us to draw an awkwardly sharp line about the time at which death begins to be bad. If we draw the line at birth, then there would be no disvalue in ending a life immediately beforehand.[9] This problem is not removed by putting the line elsewhere.

The time-relative interest account (prioritise older children over infants)

An alternative, *gradualist,* approach is to say that death starts to be bad at a certain age but then becomes worse and reaches its peak badness at some later stage.

In philosophy, this position is associated with the *time-relative interest account* (*TRIA*) of the badness of death (McMahan, 2002). The idea behind *TRIA* is that the badness of death does not consist only of the wellbeing one would have had, but also in how psychologically connected the person is to their later self. As infants are only weakly psychologically connected to their later selves, their 'interest' in living may be less than that of an adult, even though the infant has more years to live.[10] This view takes inspiration from Derek Parfit's work on personal identity and his conclusion that what prudentially matters in survival are certain kinds of psychological connectedness, rather than the physical continuity of our bodies (Parfit, 1986). *TRIA* could justify the stance taken in the earlier versions of the GBD on which saving 10-years-olds is more valuable than saving infants. It could also explain the intuitions many have that abortion is permissible - foetuses have comparatively limited interests compared to their mothers - and that earlier-stage abortions are less bad than later ones.

As we might expect, *TRIA* comes with its own theoretical challenges.

---

[9] This is more puzzling when we consider that birth is a matter of geography – when an entity moves from inside to outside the mother. Is it worse to end the life of a 35-week child outside the womb or a 36-week child inside the womb? Although we use different words to describe these - infanticide and abortion, respectively - we might wonder if there is an important moral distinction between them.

[10] As Holtug (2010) explains it, "foetuses and infants usually have rather simple psychologies and thus few of the preferences, memories and character traits they will acquire later in life".

Here's one problematic case.[11] A woman is pregnant. She can take a drug that will treat an illness but will cause substantial birth defects (a real-life example of this would be [Thalidomide](#)). According to *TRIA*, the mother would do little or nothing wrong in taking the drug because the foetus currently has quite weak interests in its own future. Intuitively, however, the mother should not take the drug.

If one accepts some sort of gradualist view, the further question is: how does it work, exactly?[12] As we saw in Figure 1 earlier, GiveWell - like the earlier version of the GBD - adopts *a* gradualist approach to saving lives, but this is not the only possible one. How much difference might it make to choose some version of *TRIA* relative to *deprivationism*?

We've illustrated some simple versions of *TRIA* in Figure 2 below, where psychological connectedness begins six months after conception and constantly increases until the 'age of full connectivity' occurs at 5, 15, or 25, after which it stays at its maximum level. In equation form, we can present the badness of death according to *TRIA* as:

wellbeing lost = wellbeing level x *years of life lost* x *TRIA discount(age of death, age of connectivity)*

Figure 2: The badness of death at different ages on *deprivationism* and versions of *TRIA*

---
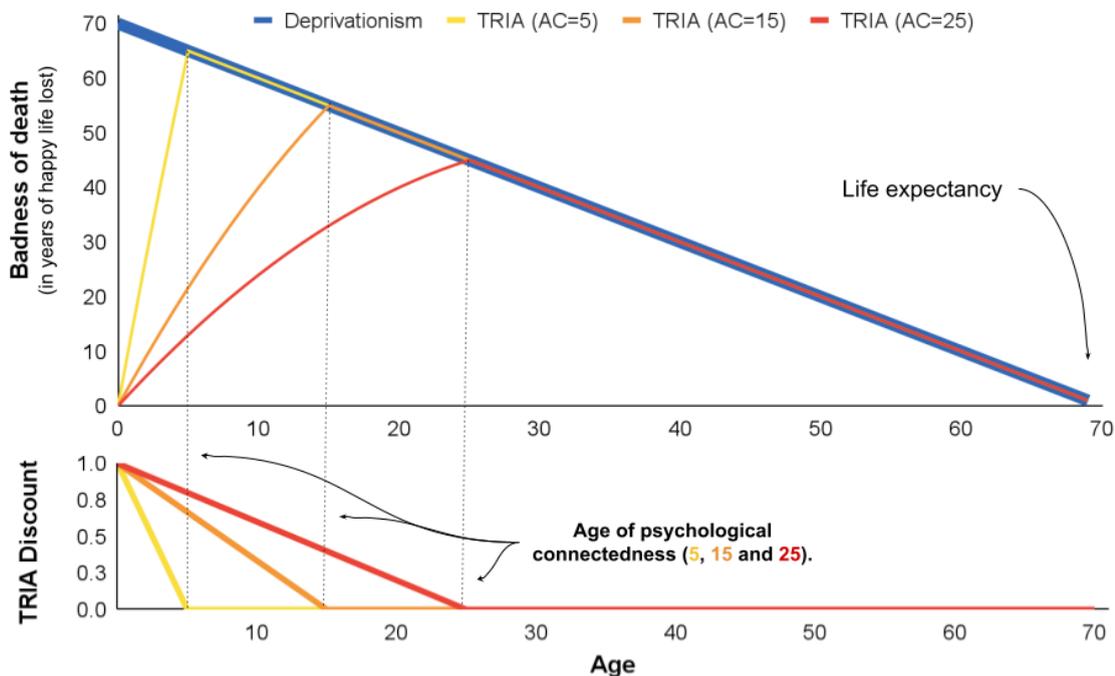
[11] For some others, see the edition collected by Gamlud and Solberg ([2019](#)), especially the chapter by Hilary Greaves.

[12] Specifically: How exactly should we model the gradually-increasing value of a life? When does someone develop an 'interest' in continuing to exist? How does this interest grow? Does it increase constantly, or does it vary with stages of development? When does it reach its peak?

As you can see, the value of saving a 5-year-old who would live 65 more years is worth 65 'happy years' if the 'age of connectivity' (AC) is 5. However, it's only worth 21 years (2/3rds smaller) if AC is 15, and worth 12 years (over 80% smaller) if AC is 25. Hence, there can be quite substantial differences in how to value a life, even within a gradualist approach. We do not aim to settle the question of which version of this view is best here, so much as notice their importance. For a discussion of what the view should be, see Norheim ([2019](#)), Mogensen ([2019](#)), and McMahan ([2019](#)).

Epicureanism (prioritise living well over living long)

The third account of the badness of death is *Epicureanism*, named after the ancient Greek philosopher Epicurus, who argued that death could not be bad for those who die on the grounds that nothing can be bad for us if we do not exist ([Hicks, 2009](#)). To be clear, Epicureans accept that dying can be painful and that a person's death can be bad for others; they only deny that a person's death can be bad *for them*. This view may seem implausible, given that we tend to think our own deaths would be bad for us, but Epicureans can raise three non-trivial puzzles for conventional views ([Solberg, 2019](#)):

- *Experience*: We often think that something cannot be good or bad for us unless we experience it. But, when we are dead, we cannot experience anything, so how can death be worse (or better) for us?

- *Time*: For something to be good(/bad) for us, there must be a time at which it is good(/bad) for us. But our future death is not bad for us whilst we are alive, and it is unclear how it is bad for us once we cease to exist.

- *Symmetry*: We don't think that our non-existence before our lives harmed us, but there is no difference between not existing before and after our lives. Therefore, our after-life non-existence is not a harm either.

Philosophers disagree on whether and how these puzzles can be solved.[13] Although what we might call *strong Epicureanism*, the view that death is not bad at all, has few takers, there may be more sympathy for *weak Epicureanism*, where death can be bad, but relatively more weight is given to living well than living long.[14] In our later calculations, we only model *strong Epicureanism*.

---

[13] For articles sympathetic to *Epicureanism*, see Rosenbaum ([1986](#)), Smuts ([2012](#)), and Hetherington ([2013](#)). For critical perspectives, see Nagel, ([1970](#)), Johansson ([2012](#)), and Bramble ([2014](#)). See Solberg ([2019](#)) for a review and further references.

[14] As the army sergeants used to say in the drill training of one of the authors: "I'm here for a good time, not a long time".

The three views discussed in this section give very different accounts of the badness of death. The first two of these, *deprivationism* and *TRIA*, are illustrated in Figure 2 above. *Epicureanism* requires no illustration as there is no badness to death at any age. We will examine the implications of these views for charity evaluations in section 3.

2.2 The neutral point

In this section, we turn to another important philosophical choice, the location of the neutral point, the level of wellbeing that existence is neither good nor bad for someone. Why does this matter?

To compare life-improvements and life-extensions, we need to combine the quality and quantity of life into a single unit. We discussed how this can be done using QALYs and DALYs at the start of section 2: on the QALY, health states are usually between 1 (full health) and 0 (equivalent to death). The reason we need a measure of quality is because we're not just interested in *how long* an intervention benefits someone, but also *how much* it benefits them. Any measure of the quality of life must include a *neutral point*. We need this to know how much better life is than some principled zero level. If we don't know how much benefit people get from different outcomes, it's impossible to allocate limited resources to do the most good. We discuss the idea that we should ignore quality of life when saving lives in a footnote.[15]

An initial issue with the neutral point is whether it even makes sense to say that lives can be bad. The QALY and DALY are both constructed by asking members of the public to make hypothetical

---

[15] Some are tempted to argue that we should give the same value to saving lives, regardless of the wellbeing those lives would have. See Open Philanthropy (2021, 5.2) and Plant (2022) for a reply, which we recapitulate here.

Although it's appealing to say that extending a life has the same value, whoever's life it is, the problem is making this consistent with other beliefs we are likely to have. If we think it's better to save someone in their 20s than in their 70s, then we already think it's correct to account for the quantity of life. Presumably, we also agree it's better to cause a bigger increase in quality of life than a small one. So, how can it be consistent to value quality when improving lives, and quantity when extending lives, but not quality when extending lives?

The QALY and DALY frameworks combine quantity and quality into a single figure because the aim is to cause the largest benefit with the available resources. What's more, if we ignore quality of life when valuing the extension of life, that will lead to other problems. If something extends someone's life *and* improves their quality of life, e.g. a surgery to remove a painful tumour, what is the value of that? Do we ignore the improvement to quality of life? How would we compare the value of that to a similar surgery, that only improved quality of life?

Another important point to make is that, in saying it's better to save one life than another, we are not saying one person has greater *intrinsic worth* than another. Rather, the claim is that the people have the same intrinsic worth, but one gets a greater benefit than the other, and that's what matters. It's better to alleviate two days of pain for Alice than one day for Bob because Alice and Bob have the same moral worth, but Alice benefits more.

trade-offs (e.g. how many years with illness X would be as good as Y years of full health?). One difference that *is* important, however, is that the QALY allows individuals to be in 'states worse than death' (SWD), whereas, on the DALY, death is the worst possible state ([Gold et al., 2002](); [Weinstein et al., 2009]()).

What's puzzling about the DALY approach is that we typically think it *is* possible to be in a state worse than death - a life of unrelenting misery and torture, for instance.[16] Part of the argument for legalising physician-assisted suicide is that death can, sometimes, be better for somebody. If living longer is always better, such an intervention could never be justified on cost-effectiveness grounds. By artificially inserting a 'floor' on how bad life can be, the DALY approach biases healthcare resources towards extending lives over improving lives.

Yet, what's unsettling about the QALY approach is this implication: if people are in states worse than death, then it is bad to extend their lives, even if those people want to keep living and we could extend their lives for free. For this reason, Scheider ([2021]()) considers the QALY objectionably 'ableist'.

There are a couple of reasons why we might find it unsettling to say that saving lives is bad. Once we recognise those, it is easier to believe there are states worse than death.[17]

First, it seems like a double unfairness if we think that it's a bad thing to provide life-extending support to unhappy lives: not only has the person been unlucky enough to end up in an unpleasant condition, but now, because of that, they won't get life-extending medical care. It's worth noting, however, that it would still count as valuable to provide life-*improving* care to those in 'states worse than death', so the QALY approach does not flatly rule out doing anything. Indeed, we often do have the intuition, at least for those who are suffering and at the end of life, that the priority is palliative care - making them comfortable rather than prolonging their lives at any cost.

Another explanation is that we don't trust the accuracy of health measures. After all, these are typically created by asking members of the public to make hypothetical trade-offs, not by asking the people in specific health states how they feel about their own lives. For instance, Bernfort et al. ([2018](), looking at a cohort of elderly Swedes, find that although 1.8% are deemed to be in states worse than death (because they have health conditions assigned negative QALY weights), 45% of

---

[16] In a survey of 667 Irish respondents, 67% disagreed with the statement "If I were severely ill with no hope of recovery, I would want to be kept alive at all costs" ([McCarthy et al., 2010]()). Out of 1,081 doctors surveyed, 88% said they do not wish to be resuscitated if their heart stops ([Periyakoil et al., 2013]()).

[17] Another is to draw a distinction between what we ought to do and what's good. Perhaps we ought to save people's lives if we can do so costlessly, even if doing so would be bad for them. A further question would be how to balance this duty against our duty to do good.

the people in that category said they were 'happy' or 'fairly happy' (55% said they were 'not happy'). Perhaps the problem is not with the idea that there can, *in theory*, be states worse than death; rather, the issue is that the QALY health measure does not, *in practice*, do a good job of drawing the line in the right place.[18]

An obvious solution is to use *subjective wellbeing* surveys, individuals' self-assessments of their happiness and/or life satisfaction, to work out how well or badly their lives are going. At the Happier Lives Institute, we have long argued for the use of subjective wellbeing measures, and for just this sort of reason (McGuire et al., 2022a; Plant, 2018). MacAskill (2022, Chapter 9) has recently argued for them too.[19]

However, this brings us to a further challenge. It's unclear where to put the neutral point on different subjective wellbeing scales, and almost no research has been done on this topic (Donaldson et al., 2020; MacAskill, 2022). On life satisfaction scales, where 0 is typically marked as 'not all satisfied' and 10 as 'completely satisfied', there is no clear indication of what respondents should interpret as the neutral point.

Some wellbeing researchers are inclined to interpret 0 as the neutral point (Layard et al., 2020), but it's unclear what the justification is for this. It runs into the same issue raised above for DALYs: it implies, implausibly, that people can only live good lives. Those who give themselves the lowest scores, 0 out of 10, would be assumed to have neutral lives.[20]

Other wellbeing researchers, such as Diener et al. (2018), appear to treat the midway point on the scale as the neutral point (i.e., 5 on a 0-10 scale). This might seem the more natural interpretation, but this also yields odd results. It suggests that a large share of the world's population has negative wellbeing (average life satisfaction in Rwanda is around 3, for example). If we think the average Rwandan lives an overall good life, that might suggest: (1) the neutral point is below 5/10 on the life satisfaction scale; (2) life satisfaction is the wrong measure of wellbeing - perhaps we believe happiness is what matters, and Rwandans are happy, but not satisfied with life[21]; (3) wellbeing will

---

[18] If we genuinely believe someone has negative wellbeing, then it may seem appropriate, in line with the standard medical principle of triage, to prioritise extending the lives of those with positive wellbeing instead. After all, they will benefit more.

[19] In response to the question of how to assess whether people's lives are overall positive or negative, he concludes, "the answer is to rely primarily on people's self-reports".

[20] It seems the main advantage of placing the neutral point at 0/10 is that it avoids giving too little to extending lives, but we also want to avoid giving too much value to extending lives.

[21] Indeed, 90% of Rwandans attest that they are 'happy' or 'very happy' (OWID, 2022). Unfortunately, the data on happiness is much sparser than life satisfaction data.

increase in Rwanda, so saving the average life is positive *in expectation*, even if it isn't right now; or (4) some combination of (1)-(3).

We might suppose, then, that the neutral point on the life satisfaction scale is *somewhere* between 0 and 5.[22] Yet, any particular choice here seems arbitrary - why 2 rather than 3, for instance? It also implies a 'lop-sided' scale where the negative end of the scale is shorter than the positive end. This clashes with the intuition that the worst possible lives are worse than the best possible lives are good. See Donaldson et al. ([2020](#)) for more details.

Empirical work on how individuals interpret the scale could be helpful but is extremely limited. A small (n = 75) survey in the UK found that respondents would choose death over life at a life satisfaction level of about 2/10 (Peasgood et al., unpublished, as referenced in [Krekel & Frijters, 2021](#)). A survey of people living in poverty in Ghana and Kenya estimated the neutral point as 0.56 ([IDinsight, 2019](#), p. 92; n = 70). There are also preliminary results from a sample of 600 in the USA, Brazil, and China that finds a neutral point of 25/100 (Jamison & Shukla, private communication). At the Happier Lives Institute, we are currently working on our own survey to explore this topic further and hope to share our results soon.

In section 3, we do not take a position on where the neutral point is when making our calculations. Instead, we explore how different neutral points between 0 and 5 on a 0-10 scale affect the cost-effectiveness of life-improving and life-extending interventions.

How much difference could the location of the neutral point make if we use subjective wellbeing scales? We will say that 1 WELLBY (wellbeing-adjusted life year) is equivalent to a 1-point *improvement* on a 0-10 life satisfaction scale for 1 year. This is similar to how 1 year of life at full health is 1 QALY, except that wellbeing runs on a 0-10 scale, not a 0-1 scale. The location of the neutral point is irrelevant for improving *quality* of life: if Alice's quality of life increases by 1 point for 2 years that is worth 2 WELLBYs. However, suppose that we can *extend* Bob's life for 1 year and Bob rates his wellbeing at 6/10. If the neutral point were 5, this is worth 1 WELLBY. If the neutral point were 0, this is worth 6 WELLBYs. Clearly, this is a large difference.

This issue becomes sharper if our efforts to save lives focus on low-income countries where, as a consequence of lower prosperity, wellbeing levels are lower. Such individuals will be closer to the neutral point than those in high-income countries, and a non-trivial proportion may be below it. Although this thought is unpleasant it should not be surprising: if we aim to help those in poverty

---

[22] If 10 is defined as 'completely satisfied', it could be argued that 10 is the neutral point and that any life containing dissatisfaction should be considered as negative. See Benatar ([2016](#)) and Vinding ([2020](#)) for further discussion.

because we think poverty is very bad, then we must already accept that those in poverty have lower wellbeing.

When we're assessing the value of extending a life in terms of wellbeing, we need to consider the change in *net* wellbeing. Take our formula for *deprivationism*:

$$\textit{wellbeing lost} = \textit{wellbeing level} \text{ x } \textit{years of life lost}$$

The existence of a neutral point implies that the wellbeing level refers to *net* wellbeing, that is, how much higher (or lower) wellbeing is than the neutral point (wherever it is).

3. How do philosophical assumptions change the cost-effectiveness of top charities?

In section 2 we discussed how philosophy influences the value of extending a life. In this section, we crunch the numbers to illustrate how much difference this could make. To do this, we compare three charities which operate in low-income countries and are considered to be some of the most cost-effective ways to help others. Two are life-improving charities: GiveDirectly, which provides cash transfers, and StrongMinds, which treats depression using group psychotherapy. The third, the Against Malaria Foundation (AMF), provides long-lasting insecticide-treated nets for protection against malaria; the main benefit of this is taken to be life-saving.[23]

We assess the cost-effectiveness of each charity using a subjective wellbeing (WELLBY) framework. We've argued that effectiveness should be assessed in WELLBYs in general (McGuire et al., 2022a). In this case, it's unclear how else we could capture the benefits of cash transfers, bednets, and psychotherapy in any other common unit. What would we use? Wealth? Health? Effects on wealth might be a good way to assess the impact of cash transfers, but not those of being alive or depression-free. Equally, effects on health might capture the value of bednets and therapy, but presumably, the value of alleviating poverty isn't just that it improves health.

To make this comparison, we combine and extend several earlier pieces of work. We've already estimated the cost-effectiveness of GiveDirectly and StrongMinds in WELLBYs. First, we undertook meta-analytic cost-effectiveness assessments of cash transfers and psychotherapy (McGuire & Plant, 2021a, McGuire & Plant 2021b). We then used those results to estimate and compare the cost-effectiveness of GiveDirectly and StrongMinds, which we consider as 'best in

---

[23] Malaria is a disease due to Plasmodium parasites transmitted by infected female Anopheles mosquitoes. In 2020, there were an estimated 627,000 deaths due to malaria (WHO, 2021).

class' organisations for implementing each intervention (see [McGuire & Plant, 2021c](#) for the original comparison and [McGuire et al., 2022b](#) for an updated version which incorporates household spillover effects). We found that GiveDirectly's cash transfers produce 8 WELLBYs/$1,000 and StrongMinds' psychotherapy produces 77 WELLBYs/$1,000, making the latter about 10 times more cost-effective than GiveDirectly.[24]

We've also done an initial estimate of the relative values of doubling consumption and saving the life of a 5-year-old using a WELLBY approach ([Donaldson et al., 2020](#)). The main two additions in this report are: (1) we estimate cost as well as effectiveness; (2) we show how the results change under several different assumptions about the badness of death and the location of the neutral point.[25] In other words, now we've got some key pieces of the jigsaw puzzle we're working out, for the first time, how they fit together.

There are two key findings.

First, the philosophical factors can radically alter the cost-effectiveness of life-extending interventions. On the assumptions most favourable to extending lives, AMF is about 30% more cost-effective than StrongMinds.[26] On the assumptions least favourable to extending lives, StrongMinds is around 12 times more cost-effective than AMF.[27] We unpack the details of this below. We consider these results noteworthy, and hopefully a useful corrective to the idea that philosophy is not an important factor in these choices.[28]

Second, these comparisons could be sensitive to the estimates of AMF's life-*improving* effects. The two we consider are: (1) averting the grief of bereaved family and friends; and (2) an income-increasing effect: a reduced incidence of malaria, particularly in childhood, may help people earn more in later life. (1) is unproblematic; we estimate this to be 2.5 WELLBYs/$1,000, a non-trivial but relatively small factor. (2) is more challenging. Taking GiveWell's income-increasing effects at face value, and converting them into WELLBYs, implies that (2) is about 17

---

[24] Our previous results ([McGuire et al., 2022b](#)) are in standard deviation changes over time (SD-years) of subjective wellbeing gained. Since these effects are standardised by dividing the raw effect by its SD, we convert it into life satisfaction points by unstandardising it with the global SD (2.2, see [row 8](#)) for life satisfaction ([Our World in Data](#)). Crucially, we assume a one-to-one exchange rate between a 1 SD change in affective mental health and subjective wellbeing measures. We're concerned this may not be justified, but our investigations so far have not supported a different exchange rate.

[25] Previously, we compared deprivationism to one version of TRIA and assumed the neutral point was around 2.5/10.

[26] Assuming *deprivationism* **and** a neutral point of 0/10

[27] Assuming *Epicureanism* **or** a neutral point of 5/10.

[28] Cf the scepticism expressed by Open Philanthropy staff [here](#) and [here](#) about how much difference philosophy makes.

WELLBYs/$1,000. This seems implausibly large. It implies that AMF (if we only count its income effect is 3x more cost-effective than GiveDirectly's cash transfers, which directly aim to reduce poverty. After further analysis, we think a more realistic figure of the income effects of bednets is 4 WELLBY/$1,000, about a quarter of the size.[29] We explain how we reached this figure in [Appendix A](#).

We will now show how we estimated the cost-effectiveness of AMF in WELLBYs, how this changes under different philosophical assumptions, and how it compares to GiveDirectly and StrongMinds. In the following three subsections, we consider the three different accounts of the badness of death and, where relevant, vary the neutral point.

3.1 Deprivationism results

We start with the simplest account, *deprivationism*. On this view:

$$badness\ of\ death = net\ wellbeing\ level \times years\ of\ life\ lost$$

We assume that the average age of the individual who dies from malaria is 20 years old[30], they would expect to live to 70, and so preventing their death leads to 50 extra years.[31] We estimate their average expected life satisfaction to be 4.95/10.[32] Hence, the WELLBYs gained by the person whose death is prevented is (4.95 - neutral point) * (70 - 20).

---

[29] This reduction is the result of incorporating all the discounts we propose to AMF's income effects in our recent essay critiquing GiveWell's CEAs ([McGuire et al., 2022c](#); [row 35, "Inputs" tab](#)).

[30] See row 18 of our [copy of GiveWell's data](#) regarding the age of death for malaria victims.

[31] To estimate the expected lifespan of malaria victims if they didn't die, we use data from the World Bank ([2020](#)). While the average lifespan for the countries AMF operates in is 59 years ([row 252](#); [World Bank, 2020](#)), UN projections expect this to increase by 11 years in the next 40 years. Therefore, we expect the average lifespan to increase to 70.

[32] In the countries that AMF operates in, the average life satisfaction is 4.5/10 ([row 251](#)). We speculate that those typically surveyed are in the upper half of income and those served by AMF are in the bottom. There's a ~0.5 life satisfaction difference in African countries between the upper and lower half ([OWID, 2020](#)). These assumptions imply that the life satisfaction of AMF's recipients is closer to 4. However, life satisfaction has been increasing in these countries in the past two decades. Using data from OWID ([2020](#)), we selected the life satisfaction data from the countries served by AMF from 2004 to 2020. We modelled life satisfaction as a linear increase over time which suggests a significant increase of 0.04 points each year (p < 0.001). To estimate the expected wellbeing for the rest of the beneficiaries' lives, we used the wellbeing we expect them to experience at the midpoint between their current age and life expectancy.

According to GiveWell, it costs $3,000 for AMF to prevent a death (on average).[33] Before we calculate the cost-effectiveness, we must also add the life-improving effects of providing malaria prevention. We estimate the grief-averting effect of preventing a death is 7 WELLBYs for each death prevented (see Appendix A.2), so 2.4 WELLBYs/$1,000. We estimate the income-increasing effects to be 4 WELLBYs/$1,000 (see Appendix A.1).

We can now calculate the cost-effectiveness of AMF under different assumptions about the location of the neutral point. If the neutral point is zero, the value of preventing a death would become 4.95 * 50 = 247 WELLBYs; this gives a cost-effectiveness of 83 WELLBYs/$1,000. This, combined with the life-improving and grief-averting benefit of 7 WELLBYs leads to a total of 90 WELLBYs,/$1,000 which is represented in the first row of Table 1 below.

This result compares favourably to the cost-effectiveness of GiveDirectly and StrongMinds, which are 8 WELLBYs/$1,000 and 77 WELLBYs/$1,000, respectively. The interpretation of this multiple is that AMF, assuming *deprivationism* and a neutral point of 0, is 10.98 times as cost-effective as GiveDirectly and 1.16 times as cost-effective as StrongMinds. We represent these as cost-effectiveness multiples in the last two columns of Table 1.

In Table 1, you can see that as the neutral point increases, the cost-effectiveness of AMF decreases considerably. At a neutral point of 5, AMF generates *negative* WELLBYs from its life-extending effects; the positive cost-effectiveness is driven entirely by AMF's life-improving effects. At a neutral point of 5, GiveDirectly is about 30% more cost-effective and StrongMinds over 10x more cost-effective than AMF. The last two rows show what happens if we take GiveWell's numbers at face value, since they estimate that malaria prevention increases income to a greater extent than we do.

---

[33] We use data from GiveWell's cost-effectiveness analysis of AMF's malaria prevention programme (see our modified copy of their spreadsheet). GiveWell estimates that AMF prevents 0.34 deaths per $1,000, (see rows 245-246; based on Pryce et al., 2018). Note that we haven't checked this number. Our results for GiveDirectly and StrongMinds are in 'per $1,000', so we normalise the benefits of AMF to a per $1,000 basis.

Table 1: Impact of different neutral points on the cost-effectiveness of AMF under *deprivationism*

| Philosophical factors	AMF cost-effectiveness (WELLBYs per $1,000)			Comparison (cost-effectiveness multiples)	
Neutral point of LS 0-10	Death prevented	Income improved	Grief averted	Total	AMF vs. GD	AMF vs. SM
0.00	82.94	4.25	2.43	89.63	10.98	1.16
1.00	66.21	4.25	2.43	72.89	8.93	0.94
2.00	49.47	4.25	2.43	56.15	6.88	0.73
3.00	32.73	4.25	2.43	39.41	4.83	0.51
4.00	15.99	4.25	2.43	22.68	2.78	0.29
5.00	-0.75	4.25	2.43	5.94	0.73	0.08
0.00	82.94	16.88	2.43	102.26	12.52	1.32
5.00	-0.75	16.88	2.43	18.57	2.27	0.24

Note: *A $1,000 donation to GiveDirectly (GD) produces 8 WELLBYs and a $1,000 donation to StrongMinds (SM) produces 77 WELLBYs. LS means a life-satisfaction scale.*

3.2 TRIA results

We now turn to the *time-relative interest account* (*TRIA*) of the badness of death. Recall that *TRIA* captures the intuition that the badness of death depends, in part, on how psychologically connected someone is to their later self. This is why the deaths of very young children are less of a loss than older ones.

To produce this calculation, we need to stipulate the age that someone reaches 'full connectedness'. In our model, we use a range of values between 5 and 25 years. This choice makes a lot of difference in this case because many of those who die from malaria are very young (see Figure 3 below).

Figure 3: Age at death from malaria using data from the Global Burden of Disease (2019)

Age at death from malaria
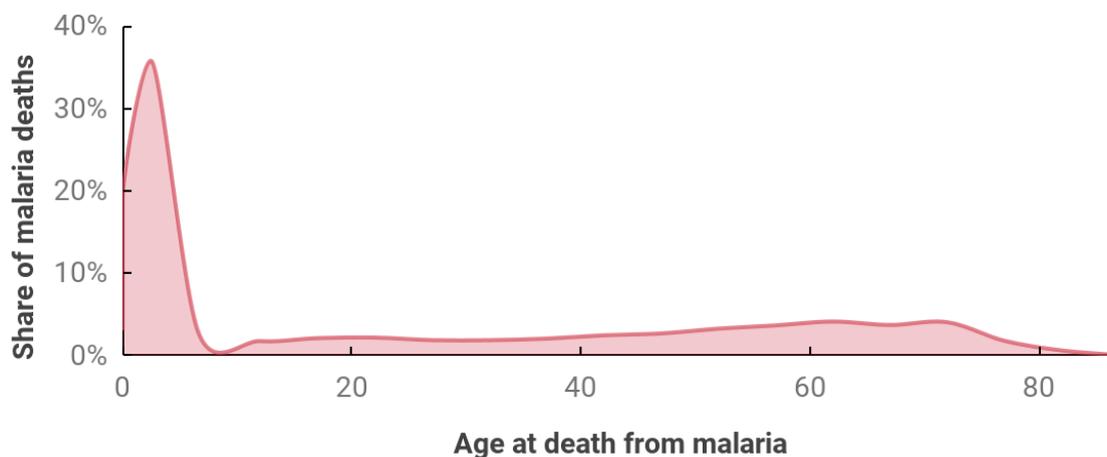
Although 58% of people who die from malaria are under five years old, due to the large number of older people who die from malaria, the average age of death is 20 (see [row 18](#)). Therefore, we split the analysis into two, one for under-5s (who we model using the average age of those in that category, 1.5 years) and over-5s (who we model using the average age of that category, 46 years).[34] Given this, we assess the loss from the over-5 deaths in the same way as *deprivationism*; only the under-5 deaths are subject to a '*TRIA* adjustment'.

We don't show all the possible combinations here, as there are too many to present. Instead, Table 2 displays combinations of the most and least generous assumptions about the neutral point and the age of connectivity, along with a middle option for both. As you can see, StrongMinds is more cost-effective than AMF in every case (between 1.1 to 12.5 times better) and GiveDirectly is less cost-effective than AMF except under the assumptions least favourable to extending lives. Again, we've added two final rows to indicate the difference it makes if we use GiveWell's larger value for life-improving effects.

Table 2: Cost-effectiveness of AMF compared to life-improving charities under TRIA

Philosophical factors			AMF cost-effectiveness (WELLBYs per $1,000)				Comparison (cost-effectiveness multiples)	
Neutral point of LS 0-10	TRIA age of connectivity	Deaths prevented	Income improved	Grief averted	Total	AMF vs. GD	AMF vs. SM
0	5	49.28	4.25	2.43	55.97	6.85	0.72
0	25	37.74	4.25	2.43	44.43	5.44	0.57
2.5	15	19.65	4.25	2.43	26.34	3.23	0.34
5	5	-0.45	4.25	2.43	6.24	0.76	0.08
5	25	-0.34	4.25	2.43	6.35	0.78	0.08
0	5	49.28	16.88	2.43	68.60	8.40	0.89
5	25	-0.34	16.88	2.43	18.98	2.32	0.25

Note: A $1,000 donation to GiveDirectly (GD) produces 8 WELLBYs and a $1,000 donation to StrongMinds (SM) produces 77 WELLBYs. LS means a life-satisfaction scale.

[34] To calculate the *TRIA* discount, we take the average age of the under-5s (1.5 years old) and weight it by the proportion of under-5 deaths (58%). Note that when we ran this analysis in R by simulating this age distribution our results did not change significantly.

3.3 Epicureanism results

For the person dying, the badness of death under *Epicureanism* is zero and, therefore, the choice of neutral point is irrelevant. For Epicureans, the effect of AMF's malaria prevention comes from its grief-averting and life-improving impact of 6.7 WELLBYs (see Table 3 below). Unsurprisingly, AMF is less cost-effective than GiveDirectly and StrongMinds in this case. As before, we've indicated how much difference it makes if the life-improving effects of malaria prevention are as large as GiveWell's estimate implies.

**Table 3**: Cost-effectiveness of AMF compared to life-improving charities under *Epicureanism*

Philosophical factors	AMF cost-effectiveness (WELLBYs per $1,000)			Comparison (cost-effectiveness multiples)	
Neutral point of LS 0-10	Deaths prevented	Income improved	Grief averted	Total	AMF vs. GD	AMF vs. SM
N/A	0	4.25	2.43	6.69	0.82	0.09
N/A	0	16.88	2.43	19.32	2.37	0.25

Note: A $1,000 donation to GiveDirectly (GD) produces 8 WELLBYs and a $1,000 donation to StrongMinds (SM) produces 77 WELLBYs. LS means a life-satisfaction scale.

3.4 Summary of results

In Figure 4, we show how the cost-effectiveness of the three charities (in WELLBYs/$1,000) changes according to the combination of philosophical views.

Figure 4: Cost-effectiveness of charities under different philosophical assumptions
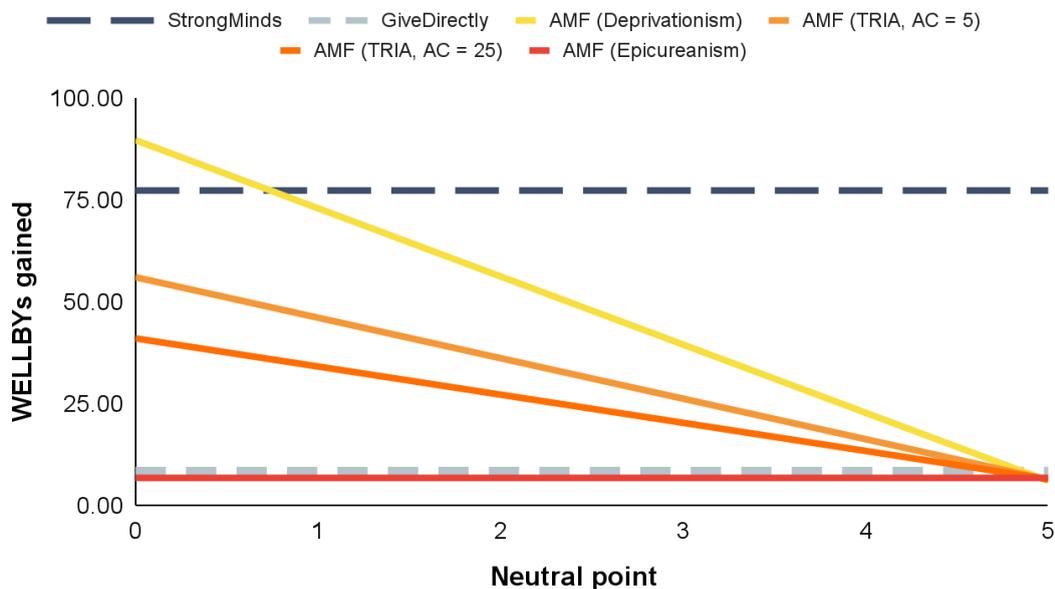
As we can see, AMF is more cost-effective than StrongMinds, but only by a small amount, and only on the assumptions most favourable to saving lives: if you accept *deprivationism* (the yellow line) and think the neutral point is less than 1. On assumptions of higher neutral points and/or different accounts of the badness of death, AMF's cost-effectiveness drops away until it becomes about 12x less cost-effective than StrongMinds and has about the same cost-effectiveness as GiveDirectly. AMF never ends up much less cost-effective than GiveDirectly on our estimates due to the non-trivial life-improving effects of bednets from averting grief and increasing later income.

4. Key uncertainties and directions for future research

In the previous section, we showed how much difference it can make if we adopt different philosophical assumptions. This raises questions for further work.

Which account of the badness of death is most plausible? We have not argued for one account in particular, so much as briefly motivated them and raised some issues. Given these are long-standing debates (again, see [Gamlund and Solberg 2019](#)), it's unclear how useful it would be for us to try to evaluate these views ourselves, rather than merely understand what the implications are of taking one or another position and communicating this to readers.

Where is the neutral point? We think the neutral point is under-theorised and under-investigated. There has only been initial empirical work (see section 2.2) and we plan to run a larger survey (n>1000) which could shed more light on how people interpret subjective scales.

How big are AMF's life-improving effects? As flagged several times, these could be quite large, and our investigation into them has only been fairly shallow. However, we're unsure how much to prioritise this, given they are a relatively smaller issue than the philosophical ones.

How should we deal with moral uncertainty? We've shown that the philosophical assumptions matter and we expect readers are unlikely to be 100% confident in a single approach. How should we account for this uncertainty? 'Moral uncertainty' is a new, important, and complicated topic at the frontiers of academic philosophy. We don't have space to get into this in any depth here, but we will sketch the three approaches that seem to be on the table. We intend to return to this and write some guidance for donors and decision-makers.

1. *My Favourite Theory*: Choose the action recommended by which moral view you have the most credence (i.e. confidence). If you had, say, the greatest credence in *deprivationism* and a neutral point of 0/10, then you should choose whatever is best given those assumptions: in this case, AMF.

2. *Maximise Expected Choiceworthiness*: Construct a weighted average of each option based on your credences in the different views and how good the option is on each view; you then choose the option that maximises expected chioceworthiness. In this case, you would award 100% of your resources to whichever life-improving or life-extending option was best. We don't run any numbers here: it should be clear from Figure 1 that the top life-improving option will be better unless one has the overwhelming weight of one's credences in both *deprivationism* and a very low neutral point.

3. *Worldview Diversification* ([Karnofksy, 2016](#)): Divide your resources across different theoretical 'buckets' in proportion to your confidence in each theory, then choose the best option for each 'bucket'. For example, if you have 30% credence in *deprivationism* and 100% credence that the neutral point is below one, you should award 30% of your resources to AMF and the rest to StrongMinds.

(1) and (2) are discussed at length in the existing philosophical literature ([Bykvist, 2017](#), [MacAskill, Bykvist, Ord 2020](#)) whereas (3) has received little attention (see [Plant, 2022b](#) and [Lloyd, 2022](#) for explorations and tentative defences).

5. Conclusion

Two key philosophical issues seem to have barely featured in discussions about whether resources do more good if we prioritise extending or improving lives. They have, however, been there all along. And they can make a big difference to cost-effectiveness calculations and our conclusions about what the overall priority is. It's reasonable to disagree about these theoretical issues, so it's essential to bring these underlying assumptions to the surface. We do not have simple answers to the problems we raise - nor do we think there are any - but we hope we have clarified them and set the stage for further work.

Appendix A: Estimating the life-improving effect of bednets

In this appendix, we explain how we estimate the SWB life-improving benefits of malaria from increased incomes for the recipients of bednets and averted grief for their loved ones.

A.1 Converting from economic benefits to WELLBYs

To estimate the effects of malaria prevention on subjective wellbeing (SWB), we would like to use direct evidence of the causal effect of malaria prevention on SWB but we found no reliable data for this.[35] The data used in some of the causal studies of malaria prevention's effect also includes SWB variables (Shih & Lin, 2019; Rawlings, 2016; Mora-García, 2018; Venkataramani, 2012). However, we have not found an SWB analysis that uses these datasets.

Instead of using SWB evidence directly, we convert from malaria prevention's estimated economic benefits to WELLBYs. GiveWell estimates that spending $1,000 on AMF's malaria prevention programme increases economic benefits by 18 log-units (or about 13 doublings of income, row 247, "Modified" tab). McGuire et al. (2022c) explains how GiveWell arrived at these figures and argues why these effects on income are likely to be smaller and less certain than GiveWell's estimates.

We translate the economic benefits into WELLBYs by estimating what a log-unit increase in economic benefits predicts in WELLBYs. To estimate this figure, we use the model in our cash transfers meta-analysis where we moderate the SWB effect of cash transfers by the relative size of the transfer compared to previous income (Model 2 in Table 2 of McGuire et al., 2022d). We calculated

---

[35] We looked at two sources of direct evidence on SWB effects as sanity checks but did not update much on them.

Two natural experiments in Ghana and Colombia (Pasha et al., 2019; Carrillo, 2020) found that increased rainfall around the time of a child's birth led to them having a higher likelihood of mental health issues later in life (20+ years later). Both papers argue from correlational evidence that this appears to be through an increase in mosquito-borne diseases (such as malaria). But neither of the studies pin down a causal link that runs through malaria prevalence. But according to Oheneba-Dornyo et al. (2022), for every 1 mm increase in rainfall, there's an additional case of malaria two months afterwards. If rainfall increases by 79 mm (the SD of rainfall in Ghana) the predicted increase in the population with malaria cases will only be 0.1%. Also, rainfall seems too weakly related to malaria to be a primary driver of this effect, so we are very uncertain about this evidence. That seems insufficient to explain the rainfall shocks as a malaria story.

The second source of evidence we considered involves cerebral malaria. In ~1% of malaria cases, the malaria attacks the brain (Idro, 2010), which can result in mental debilitation. There's correlational evidence on small sample sizes that ~1 year after being infected with malaria, children with cerebral malaria have a greater likelihood of showing signs of a behavioural disorder (odd ratio: 8.5, Christensen & Eslick, 2015) or a mental disorder when compared to a control group of healthy kids (Ssenkusu et al., 2016; Bangirana et al., 2011; Boivin et al., 2018). But due to cerebral malaria's rarity, we think that preventing this would only constitute a small benefit. Neither of these separate strands of evidence update us much about the plausibility of the SWB effect of malaria exposure in childhood.

the total benefits in SD-years using the 'area of the triangle' formula as illustrated and described in Figure 1 of McGuire and Plant (2021a). Since a 1-unit increase in log-economic benefits is approximately a 70% increase in household income, plugging a 70% increase in income (1 log-unit) into Model 2 from Table 2 (McGuire et al., 2022d) leads to a 0.6 SD-year increase in SWB (row 23, "Inputs" tab). To convert from SD-years to WELLBYs we multiply the SD-years by the average SD of life satisfaction (2.17, see row 8, "Inputs" tab), which results in 0.6 x 2.17 = 1.3 WELLBYs. If we take these numbers at face value, $1,000 donated to AMF's malaria prevention programme creates 18 x 1.3 = 23 WELLBYs due to income-increasing effects. Note that this includes household spillovers because GiveWell includes household spillovers in their figure. In GiveWell's analysis, non-recipients receive a benefit that is 25% of the benefit received by the recipient. This is equivalent to GiveWell's 2x household multiplier for households of four (row 120, "Modified" tab).

However, there's a key difference between the way that cash transfers and malaria prevention increase incomes. When it comes to increases in income, part of the SWB benefit may be due to a comparison effect - comparing yourself with others and finding yourself better or worse off (Clark et al., 2018; Easterlin, 2021). Comparison effects apply to cash transfers, but they probably do not apply to malaria prevention. With mass distribution of bednets, the benefits accrue to everyone in an area, while cash transfers are only sent to the poorest. We estimate that these comparison/relative income effects comprise 26% of the total income effect.[36] Including this adjustment discounts the life-improving effect of malaria prevention from 23 down to 17 WELLBYs.

The final step is to incorporate various further discounts to AMF's income effects that we proposed in our recent essay critiquing GiveWell's CEAs (McGuire et al., 2022c). We won't restate all those details here, but the short version is that we identified several issues with GiveWell's calculations. We estimate that, cumulatively, this leads to a 75% reduction in the income effect relative to what GiveWell estimates (row 35, sheet "Inputs"), or a (1 - 0.75) x 17 = 4 WELLBY effect. To frame this effect, we compare it to the other charities we've analysed in Figure A1 below.

---

[36] The approach we took to estimate the importance of relative income effects was to see how much the magnitude of the absolute income coefficient decreases when you introduce a relative income coefficient (comparing yourself to your peers) to the equation. We interpret this ratio of coefficients (absolute income coefficient after / absolute income coefficient before the introduction of the relative term) as the share of absolute income effects from the total income effects. This is 74% for Reyes-Garcia et al. (2015), a study whose sample seems similar to that of the beneficiaries of AMF. The average income in the AMF countries in 2021 is $1,171 (row 256, "Modified" tab), and the average income is $1,062 for Reyes-Garcia et al. (2015). Relative income effects will likely grow with total income.

Figure A1: Life-improving effect of charities in WELLBYs
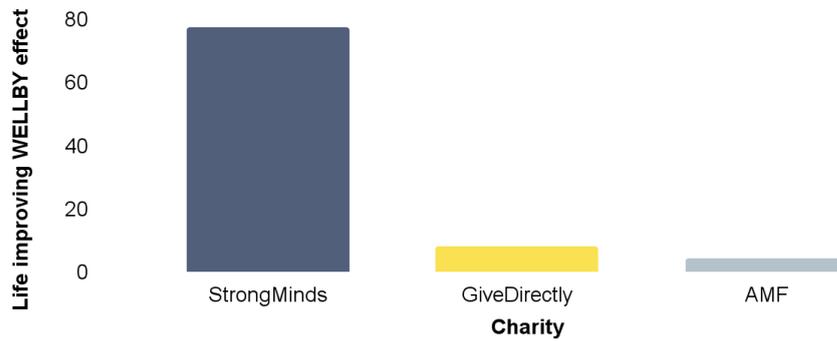
A.2 Grief spillovers

We do a shallow calculation for grief in the same way we did in Donaldson et al. (2020). The best estimate we found is from Oswald and Powdthavee (2008): a panel study in the UK which finds the effect on life satisfaction due to the death of a child in the last year as being -0.72 (adjusted for a 0-10 scale). According to Clark et al. (2018), the duration of grief is ~5 years. Based on data from the UNDP, we calculate that the average household size across the beneficiary countries (excluding the recipient of the nets) is 4.03 people (row 16). Hence, an overall effect of grief per death prevented is (0.72 x 5 x 0.5) x 4.03 = 7.26 WELLBYs. However, we think this is an upper bound because it doesn't account for the counterfactual grief averted. If you avert the death of someone, they will still die at some point in the future and the people who love them will still grieve.